



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2012

---

## **The Aquila comparison project: the effects of feedback and numerical methods on simulations of galaxy formation**

Scannapieco, C ; Wadepuhl, M ; Parry, O H ; Navarro, J F ; Jenkins, A ; Springel, V ; Teyssier, R ; Carlson, E ; Couchman, H M P ; Crain, R A ; Vecchia, C D ; Frenk, C S ; Kobayashi, C ; Monaco, P ; Murante, G ; Okamoto, T ; Quinn, T ; Schaye, J ; Stinson, G S ; Theuns, T ; Wadsley, J ; White, S D M ; Woods, R

**Abstract:** We compare the results of various cosmological gas-dynamical codes used to simulate the formation of a galaxy in the  $\Lambda$  cold dark matter structure formation paradigm. The various runs (13 in total) differ in their numerical hydrodynamical treatment [smoothed particle hydrodynamics (SPH), moving mesh and adaptive mesh refinement] but share the same initial conditions and adopt in each case their latest published model of gas cooling, star formation and feedback. Despite the common halo assembly history, we find large code-to-code variations in the stellar mass, size, morphology and gas content of the galaxy at  $z=0$ , due mainly to the different implementations of star formation and feedback. Compared with observation, most codes tend to produce an overly massive galaxy, smaller and less gas rich than typical spirals, with a massive bulge and a declining rotation curve. A stellar disc is discernible in most simulations, although its prominence varies widely from code to code. There is a well-defined trend between the effects of feedback and the severity of the disagreement with observed spirals. In general, models that are more effective at limiting the baryonic mass of the galaxy come closer to matching observed galaxy scaling laws, but often to the detriment of the disc component. Although numerical convergence is not particularly good for any of the codes, our conclusions hold at two different numerical resolutions. Some differences can also be traced to the different numerical techniques; for example, more gas seems able to cool and become available for star formation in grid-based codes than in SPH. However, this effect is small compared to the variations induced by different feedback prescriptions. We conclude that state-of-the-art simulations cannot yet uniquely predict the properties of the baryonic component of a galaxy, even when the assembly history of its host halo is fully specified. Developing feedback algorithms that can effectively regulate the mass of a galaxy without hindering the formation of high angular momentum stellar discs remains a challenge.

DOI: <https://doi.org/10.1111/j.1365-2966.2012.20993.x>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-70312>

Journal Article

Published Version

Originally published at:

Scannapieco, C; Wadepuhl, M; Parry, O H; Navarro, J F; Jenkins, A; Springel, V; Teyssier, R; Carlson, E; Couchman, H M P; Crain, R A; Vecchia, C D; Frenk, C S; Kobayashi, C; Monaco, P; Murante, G; Okamoto, T; Quinn, T; Schaye, J; Stinson, G S; Theuns, T; Wadsley, J; White, S D M; Woods, R (2012).

The Aquila comparison project: the effects of feedback and numerical methods on simulations of galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 423(2):1726-1749.  
DOI: <https://doi.org/10.1111/j.1365-2966.2012.20993.x>

# The Aquila comparison project: the effects of feedback and numerical methods on simulations of galaxy formation

C. Scannapieco,<sup>1★</sup> M. Wadepuhl,<sup>2</sup> O. H. Parry,<sup>3,4</sup> J. F. Navarro,<sup>5</sup> A. Jenkins,<sup>3</sup>  
V. Springel,<sup>6,7</sup> R. Teyssier,<sup>8,9</sup> E. Carlson,<sup>10</sup> H. M. P. Couchman,<sup>11</sup> R. A. Crain,<sup>12,13</sup>  
C. Dalla Vecchia,<sup>14</sup> C. S. Frenk,<sup>3</sup> C. Kobayashi,<sup>15,16</sup> P. Monaco,<sup>17,18</sup> G. Murante,<sup>17,19</sup>  
T. Okamoto,<sup>20</sup> T. Quinn,<sup>10</sup> J. Schaye,<sup>13</sup> G. S. Stinson,<sup>21</sup> T. Theuns,<sup>3,22</sup> J. Wadsley,<sup>11</sup>  
S. D. M. White<sup>2</sup> and R. Woods<sup>11</sup>

<sup>1</sup>Leibniz-Institute for Astrophysics Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany

<sup>2</sup>Max-Planck Institute for Astrophysics, Karl-Schwarzschild-Str. 1, 85741 Garching, Germany

<sup>3</sup>Institute for Computational Cosmology, Department of Physics, Durham University, Durham DH1 3LE

<sup>4</sup>Department of Astronomy, University of Maryland, College Park, MD 20745, USA

<sup>5</sup>Department of Physics and Astronomy, University of Victoria, Victoria, BC V8P 5C2, Canada

<sup>6</sup>Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnengasse 35, 69118 Heidelberg, Germany

<sup>7</sup>Zentrum für Astronomie der Universität Heidelberg, ARI, Mönchhofstr. 12-14, 69120 Heidelberg, Germany

<sup>8</sup>CEA, IRFU, SAp, 91191 Gif-sur-Yvette, France

<sup>9</sup>Institute for Theoretical Physics, University of Zürich, CH-8057 Zürich, Switzerland

<sup>10</sup>Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195-1580, USA

<sup>11</sup>Department of Physics and Astronomy, McMaster University, Hamilton, ON L8S 4M1, Canada

<sup>12</sup>Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia

<sup>13</sup>Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, the Netherlands

<sup>14</sup>Max Planck Institute for Extraterrestrial Physics, Gießenbachstraße 1, 85748 Garching, Germany

<sup>15</sup>School of Physics, Astronomy and Mathematics, University of Hertfordshire, Hatfield AL10 9AB

<sup>16</sup>Research School of Astronomy and Astrophysics, The Australian National University, Cotter Road, Weston, ACT 2611, Australia

<sup>17</sup>Dipartimento di Fisica – Sezione di Astronomia, Università di Trieste, via Tiepolo 11, I-34131 Trieste, Italy

<sup>18</sup>INAF, Osservatorio Astronomico di Trieste, via Tiepolo 11, I-34131 Trieste, Italy

<sup>19</sup>INAF, Osservatorio Astronomico di Torino, Strada Osservatorio 20, I-10025 Pino Torinese, Italy

<sup>20</sup>Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, 305-8577 Ibaraki, Japan

<sup>21</sup>Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany

<sup>22</sup>Universiteit Antwerpen, Campus Groenenborger, Groenenborgerlaan 171, B-2020 Antwerpen, Belgium

Accepted 2012 March 25. Received 2012 March 23; in original form 2011 November 29

## ABSTRACT

We compare the results of various cosmological gas-dynamical codes used to simulate the formation of a galaxy in the  $\Lambda$  cold dark matter structure formation paradigm. The various runs (13 in total) differ in their numerical hydrodynamical treatment [smoothed particle hydrodynamics (SPH), moving mesh and adaptive mesh refinement] but share the *same* initial conditions and adopt in each case their latest published model of gas cooling, star formation and feedback. Despite the common halo assembly history, we find large code-to-code variations in the stellar mass, size, morphology and gas content of the galaxy at  $z = 0$ , due mainly to the different implementations of star formation and feedback. Compared with observation, most codes tend to produce an overly massive galaxy, smaller and less gas rich than typical spirals, with a massive bulge and a declining rotation curve. A stellar disc is discernible in most simulations, although its prominence varies widely from code to code. There is a well-defined trend between the effects of feedback and the severity of the disagreement with observed spirals. In general, models that are more effective at limiting the baryonic mass of the galaxy

★E-mail: cscannapieco@aip.de

come closer to matching observed galaxy scaling laws, but often to the detriment of the disc component. Although numerical convergence is not particularly good for any of the codes, our conclusions hold at two different numerical resolutions. Some differences can also be traced to the different numerical techniques; for example, more gas seems able to cool and become available for star formation in grid-based codes than in SPH. However, this effect is small compared to the variations induced by different feedback prescriptions. We conclude that state-of-the-art simulations cannot yet uniquely predict the properties of the baryonic component of a galaxy, even when the assembly history of its host halo is fully specified. Developing feedback algorithms that can effectively regulate the mass of a galaxy without hindering the formation of high angular momentum stellar discs remains a challenge.

**Key words:** methods: numerical – galaxies: evolution – galaxies: formation – cosmology: theory.

## 1 INTRODUCTION

Numerical simulations play a central role in studies of cosmic structure formation. Collisionless  $N$ -body simulations have now become the main theoretical tool to predict the non-linear evolution of dark-matter-dominated structures once initial conditions are specified. Their high accuracy and huge dynamic range have allowed a detailed comparison of their outcome with observations of the large-scale structure of the Universe. The impressive agreement between these observations and the predictions of the  $\Lambda$  cold dark matter ( $\Lambda$ CDM) model has helped to establish it as the current paradigm of structure formation (Springel, Frenk & White 2006).

Simulating the evolution of the visible Universe is much more complex, because it requires understanding the many astrophysical processes which drive the evolution of the baryonic component under the gravitational influence of the dark matter. Numerical hydrodynamics in cosmological simulations has traditionally used either the Lagrangian smoothed particle hydrodynamics (SPH) technique (Gingold & Monaghan 1977; Lucy 1977; Monaghan 1992; Katz, Weinberg & Hernquist 1996; Springel 2010b) or Eulerian grid-based solvers sometimes aided by adaptive mesh refinement (AMR) techniques (Cen & Ostriker 1992; Bryan & Norman 1995; Kravtsov 1999; Fryxell et al. 2000; Teyssier 2002; Quilis 2004).

Both approaches have advantages and disadvantages. It is widely appreciated that SPH is not able to capture shocks with high accuracy and that in certain situations fluid instabilities can be suppressed, at least for standard implementations of SPH (Agertz et al. 2007; Creasey et al. 2011). On the other hand, mesh-based codes are not Galilean invariant and may in some cases generate entropy spuriously through artificial mixing (Wadsley, Veeravalli & Couchman 2008). As a result, even for some simple non-radiative problems, Lagrangian and Eulerian codes do not converge to the same solution (e.g. Okamoto et al. 2003; Agertz et al. 2007; Tasker et al. 2008; Mitchell et al. 2009). Novel techniques, such as the Lagrangian, moving mesh AREPO code introduced by Springel (2010a), hold the promise of improving this state of affairs, but their application is still in its infancy.

An even more uncertain ingredient of galaxy formation simulations is the descriptions of poorly understood physical processes such as star formation and feedback. The huge dynamic range between the super-Megaparsec scales needed to follow the hierarchical growth of a galaxy and the sub-parsec scales that govern the transformation of its gas into stars implies that direct simulation of all relevant physical processes is still out of reach of even the most

powerful computers and best available algorithms. Star formation and feedback are therefore introduced in cosmological simulations as ‘subgrid’ parametrized prescriptions of limited physical content and lacking numerical rigor.

These difficulties have hampered the progress of simulations of galaxy formation within the  $\Lambda$ CDM paradigm, but a few results of general applicability have nevertheless emerged. For example, the formation of *realistic disc galaxies* in dark matter haloes formed hierarchically, as expected in  $\Lambda$ CDM, was recognized as a challenge even in early simulations (see e.g. Navarro, Frenk & White 1995; Navarro & Steinmetz 1997). Simulated galaxies suffered from ‘overcooling’ and from a dearth of angular momentum due to the transfer of angular momentum from the baryonic component to the dark matter during the many merger events that characterize hierarchical assembly (Navarro & Benz 1991; Sommer-Larsen, Gelato & Vedel 1999; Navarro & Steinmetz 2000). Feedback, as a general heating mechanism that can prevent overcooling and regulate the assembly of a galaxy whilst avoiding catastrophic angular momentum losses, emerged as a crucial ingredient of any successful galaxy formation simulation (Navarro & Steinmetz 1997; Weil, Eke & Efsthathiou 1998; Sommer-Larsen, Gelato & Vedel 1999; Scannapieco et al. 2008; Zavala, Okamoto & Frenk 2008).

Considerable progress has been made since this time: recent simulations have shown that the angular momentum problem can indeed be alleviated when feedback from evolving stars, in particular supernovae (SNe), is included (e.g. Okamoto et al. 2005; Governato et al. 2007; Scannapieco et al. 2008, 2009; Ceverino & Klypin 2009; Joung, Cen & Bryan 2009; Colin et al. 2010; Sales et al. 2010; Stinson et al. 2010). Some authors have also investigated alternative feedback mechanisms, such as energy liberated during the formation of supermassive black holes (SMBHs) as well as that carried by cosmic rays (CRs; e.g. Di Matteo et al. 2003; Springel, Di Matteo & Hernquist 2005a; Jubelgas et al. 2008; Booth & Schaye 2009; Fabjan et al. 2010; Wadepuhl & Springel 2011).

Despite progress, difficulties remain. For example, simulations tend to allow far too many baryons to accrete into galaxies to be consistent with the observed stellar mass function of galaxies (see e.g. Guo et al. 2010). In addition, although stellar discs do form, they are often too concentrated, with steeply declining rotation curves at odds with observation (e.g. Abadi et al. 2003a,b; Stinson et al. 2010).

The difficulties in simulating disc galaxies highlighted above are compounded by the limited guidance afforded by analytic and semi-analytic models of galaxy formation. In such modelling, the



properties of disc galaxies are typically envisioned to reflect those of their surrounding haloes (Fall & Efstathiou 1980; Kauffmann, White & Guiderdoni 1993; Dalcanton, Spergel & Summers 1997; Mo, Mao & White 1998; Bower et al. 2006; De Lucia & Blaizot 2007): for example, haloes with high net spin and quiet recent merger histories are commonly assumed to be likely sites of disc galaxy formation. However, there are growing indications that these assumptions might be too simplistic. Scannapieco et al. (2009) and Stinson et al. (2010), for example, find no clear relation between the presence of discs and the spin parameter of the halo: discs form in haloes with low and high spin parameters. Moreover, Scannapieco et al. (2009) report that disc formation is not assured by a quiescent assembly history, since they can be destroyed not only by major or intermediate-mass mergers, but also by secular processes such as a misalignment between the gaseous and stellar discs.

These difficulties have led to little consensus on what determines the morphology of a galaxy, what the main feedback mechanisms are and what role they play on different mass scales and at different times. Indeed, there is even debate about whether the difficulties in reproducing realistic discs are predominantly a consequence of insufficient numerical resolution (e.g. Governato et al. 2004), inappropriate modelling of the relevant physics (e.g. Mayer, Governato & Kaufmann 2008; Piontek & Steinmetz 2011) or a failure of the cosmological model (e.g. Sommer-Larsen & Dolgov 2001).

Progress in this unsettled field requires at least a careful evaluation of the different numerical techniques, resolution and choice of subgrid physics adopted by various authors. Most groups do carry out and publish resolution tests and convergence studies of their own numerical set-up. However, because of the complexity of the problem, the lack of telling test cases with known solution, and the absence of clear theoretical predictions for individual systems, each group chooses to tune the relevant numerical parameters according to different priorities and/or prejudice, and there has so far been little effort invested in comparing the results of different techniques and codes. Would they give similar results if they followed the formation of a galaxy in the *same* dark matter halo?

The main goal of the Aquila project is to address this question by comparing the predictions of different codes using *common initial conditions* and a *homogeneous set of analysis tools*. Rather than focusing on whether individual codes perform better or worse than others, we contrast their predictions for the stellar mass, angular momentum content, star formation rates (SFRs) and galaxy size, with observation.

This paper is organized as follows. Section 2 describes the initial conditions and the simulation set-up. Section 3 compares the galaxy morphology, star formation history, size, angular momentum, and gas content of the 13 simulated galaxies, together with a brief discussion of the effects of numerical resolution on the results. Finally, Section 4 summarizes our main findings and lists our main conclusions.

## 2 THE SIMULATIONS

### 2.1 The codes

The Aquila project consists of 13 different gas-dynamical simulations of the formation of a galaxy in a  $\Lambda$ CDM halo of similar mass to that of the Milky Way. Nine different codes were used for the project (two codes were run three times each with different subgrid physics modules). The various codes differ in their hydrodynamical technique (SPH, AMR, moving mesh): seven codes use the SPH technique, six of which are based on GADGET (hereafter G3 for short;

see Springel 2005) and one on GASOLINE (hereafter referred to as GAS; Wadsley, Stadel & Quinn 2004).

The G3-based codes share the same numerical gravity/hydrodynamical treatment but differ in their cooling/star formation/feedback modules. G3 refers to the standard Springel & Hernquist (2003) implementation; G3-CS refers to the code presented in Scannapieco et al. (2005, 2006); G3-TO to that developed by Okamoto et al. (2005, 2010) and Okamoto, Nemmen & Bower (2008); G3-GIMIC is described in Crain et al. (2009); G3-MM is introduced by Murante et al. (2010) and G3-CK by Kobayashi et al. (2007). Of the two codes that do not use SPH one is the RAMSES AMR code (hereafter R, for short; see Teyssier 2002), and the other is the moving mesh code AREPO (Springel 2010a).

Each code was run by the group responsible for its development, adopting (independently from the choices made by other Aquila participants) their latest published model of cooling, star formation and feedback. These differ from code to code. Regarding radiative cooling, for example, some codes assume primordial abundances to compute cooling rates; others use metal-dependent cooling rate tables and in some cases cooling is implemented on an element-by-element basis. Star formation also varies from code to code, although in nearly all cases the efficiency of transformation of gas into stars is set by attempting to reproduce the Kennicutt–Schmidt empirical relation (Kennicutt 1998) in simulations of isolated discs (see also Fig. A1).

The numerical treatment of feedback also varies from code to code. In most cases, *thermal feedback* is used, where SN energy is injected into the interstellar medium (ISM) as thermal energy. The dispersal of the input energy is usually delayed artificially, in order to promote the pressurization of the ISM, the onset of winds and the effective regulation of subsequent star formation. A few codes adopt *kinetic feedback*, where kinetic energy is dumped directly into the gas. In the G3-TO code, the outflowing gas is temporarily *decoupled* hydrodynamically from the rest of the ISM to ensure that the specified mass loading (wind mass per unit mass of stars formed) and velocity are not modified by viscous drag until gas escapes the star-forming region.

Two further simulations were run with the standard G3 code: one (G3-BH) that included, in addition to SN, the feedback energy associated with the assembly of SMBHs and a third (G3-CR) where another form of feedback, that associated with energy deposition by CRs, was also included.

Two further RAMSES runs are also part of our series, one (R-LSFE) where the star formation time-scale is much longer than the fiducial choice, delaying substantially the transformation of gas into stars, and another (R-AGN) where the feedback energy was augmented by the contribution of a putative AGN associated with a central SMBH.

The main characteristics of each code are summarized in Table 1. Appendix A presents a more detailed description of each of the codes and their numerical choices.

### 2.2 Initial conditions

All simulations share *the same initial conditions* (ICs), a zoomed-in resimulation of one of the haloes of the Aquarius Project (halo ‘Aq-C’, in the notation of Springel et al. 2008). The ICs were generated using Fourier methods as described in Springel et al. (2008), modified to include a gas component. The displacement field is first calculated on a set of grids and then interpolated on to the nodes of the unperturbed particle positions, chosen from a glass-like configuration. For SPH codes these particles represent

**Table 1.** Summary of code characteristics and subgrid physics.

Code	Reference	Type	UV background ( $z_{UV}$ )	(spectrum)	Cooling	Feedback
G3 (GADGET3)	[1]	SPH	6	[10]	Primordial [13]	SN (thermal)
G3-BH	[1]	SPH	6	[10]	Primordial [13]	SN (thermal), BH
G3-CR	[1]	SPH	6	[10]	Primordial [13]	SN (thermal), BH, CR
G3-CS	[2]	SPH	6	[10]	Metal dependent [14]	SN (thermal)
G3-TO	[3]	SPH	9	[11]	Element-by-element [15]	SN (thermal+kinetic)
G3-GIMIC	[4]	SPH	9	[11]	Element-by-element [15]	SN (kinetic)
G3-MM	[5]	SPH	6	[10]	Primordial [13]	SN (thermal)
G3-CK	[6]	SPH	6	[10]	Metal dependent [14]	SN (thermal)
GAS (GASOLINE)	[7]	SPH	10	[12]	Metal dependent [16]	SN (thermal)
R (RAMSES)	[8]	AMR	12	[10]	Metal dependent [14]	SN (thermal)
R-LSFE	[8]	AMR	12	[10]	Metal dependent [14]	SN (thermal)
R-AGN	[8]	AMR	12	[10]	Metal dependent [14]	SN (thermal), BH
AREPO	[9]	Moving mesh	6	[10]	Primordial [13]	SN (thermal)

Note: [1] Springel et al. (2008); [2] Scannapieco et al. (2005), Scannapieco et al. (2006); [3] Okamoto et al. (2010); [4] Crain et al. (2009); [5] Murante et al. (2010); [6] Kobayashi, Springel & White (2007); [7] Stinson et al. (2006); [8] Teyssier (2002), Rasera & Teyssier (2006), Dubois & Teyssier (2008); [9] Springel (2010a); [10] Haardt & Madau (1996); [11] Haardt & Madau (2001); [12] Haardt & Madau (private communication); [13] Katz et al. (1996); [14] Sutherland & Dopita (1993); [15] Wiersma, Schaye & Smith (2009a); [16] Shen, Wadsley & Stinson (2010).

the matter distribution, while for the AMR initial conditions they represent the dark matter only.

The displacement field is used to perturb the particle positions and to assign them velocities consistent with the growing mode of the density fluctuations. For AMR runs the density and velocity fields of the gas are needed on a set of meshes. These quantities were calculated in the same way as described above except that the displacement and density fields were interpolated on to the vertices of a pre-defined set of nested hierarchical grids tailored to requirements of the AMR code.

For SPH runs, each high-resolution dark matter particle is split into two to create one dark matter and one gas particle, with relative masses given by the assumed value of the universal baryon abun-

dance parameter  $\Omega_b$  (Table 2). The positions of the new particles are such that their centre of mass position and velocity are identical to those of the original particle.

The selected halo, Aq-C, has a present-day mass similar to the Milky Way ( $\sim 1.6 \times 10^{12} M_\odot$ ; e.g. Dehnen, McLaughlin & Sachania 2006; Smith et al. 2007; Li & White 2008; Xue et al. 2008; Watkins, Evans & An 2010) and has a relatively quiet formation history. It is also mildly isolated at  $z = 0$ , with no neighbouring halo more massive than half its mass within a radius of  $1 h^{-1} \text{ Mpc}$ . Maps of the dark matter distribution in boxes of various sizes are shown in Fig. 1.

### 2.3 Cosmology

We assume a  $\Lambda$ CDM cosmology with the following parameters:  $\Omega_m = 0.25$ ,  $\Omega_\Lambda = 0.75$ ,  $\sigma_8 = 0.9$ ,  $n_s = 1$  and a Hubble constant of  $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1} = 73 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . These parameters are consistent with the *Wilkinson Microwave Anisotropy Probe* (WMAP) 1- and 5-year results at the  $3\sigma$  level and are identical to the parameters used for the Millennium and Millennium-II simulations (Springel et al. 2005b; Boylan-Kolchin et al. 2009). The value of  $\Omega_b$  used in each simulation is given in Table 2.

The Millennium-II is, in fact, a resimulation of the cosmological volume from which the Aquarius haloes were originally selected.<sup>1</sup> Aq-C is thus present both in this simulation and in the semi-analytic model of Guo et al. (2011) which was tuned to fit the luminosity, stellar mass, size and gas content functions measured for galaxies in the Sloan Digital Sky Survey (SDSS). Similarly, Cooper et al. (2010) used the GALFORM code to model star formation in all six Aquarius haloes,<sup>2</sup> using parameters very similar to those of Bower et al. (2006), which reproduce local galaxy luminosity functions. The properties found for the central galaxy of Aq-C in these two models thus give an indication of what direct simulations should produce if implementation on a cosmological volume is to reproduce observed galaxy abundances.

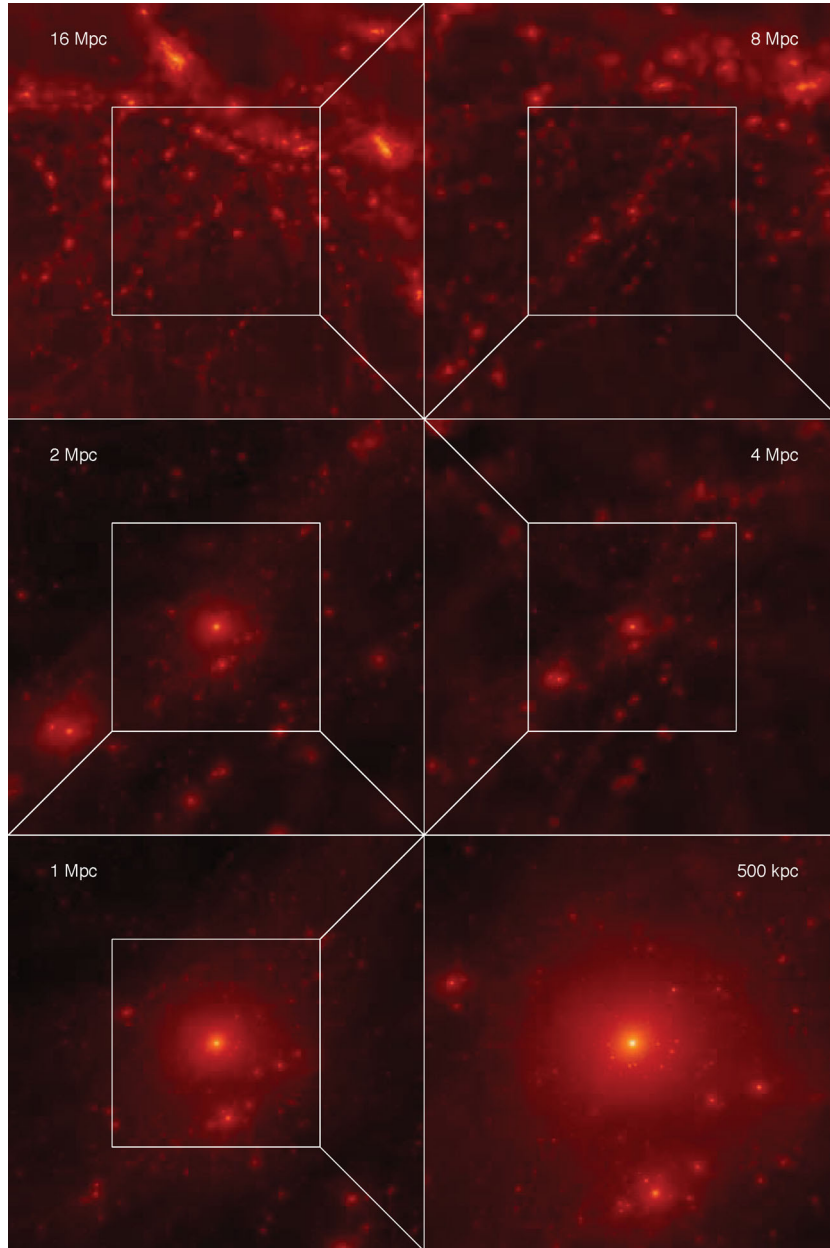
<sup>1</sup> In the convention of the Aquarius Project (lower numbers indicate higher resolution), the Millennium-II simulation has a resolution intermediate between levels 5 and 6 (the two resolutions used in our set of simulations).

<sup>2</sup> In this case, the level-4 resolution simulations were used.

**Table 2.** Code parameters for simulations at level-5 resolution (level-6 parameters are given in parentheses).

Code	$f_b$ ( $\Omega_b/\Omega_m$ )	$m_{DM}$ ( $10^6 M_\odot$ )	$m_{gas}$ ( $10^6 M_\odot$ )	Softening $\epsilon_g^{z=0}$ (kpc)	$z_{fix}$
G3					
G3-BH					
G3-CR	0.16	2.2	0.4	0.7	0
G3-CS		(17)	(3.3)	(1.4)	(0)
G3-CK					
AREPO					
G3-TO	0.18	2.1	0.5	0.5	3
G3-GIMIC		(17)	(3.7)	(1)	(3)
G3-MM	0.16	2.2	0.4	0.7	2
		(17)	(3.3)	(1.4)	(2)
GAS	0.18	2.1	0.5	0.46	8
		(17)	(3.7)	(0.9)	(8)
R	0.16	1.4	0.2	0.26	9
R-LSFE		(11)	(1.8)	(0.5)	(9)
R-AGN					

Note:  $f_b$ : baryon fraction;  $m_{DM}$ : mass of dark matter particles in the high-resolution region;  $m_{gas}$ : initial mass of gas particles;  $\epsilon_g^{z=0}$ : gravitational softening at  $z = 0$ ;  $z_{fix}$ : redshift after which the gravitational softening is kept fixed in physical coordinates. The softening is fixed in comoving coordinates at  $z > z_{fix}$  (see Appendix B).



**Figure 1.** Maps of the dark matter distribution in the region surrounding the Aquila halo at  $z = 0$ . The dark matter is projected in cubic volumes of side length as indicated in each panel. Pixel brightness corresponds to the dark matter density using a logarithmic scale.

## 2.4 The runs

All 13 simulations were run at two different numerical resolutions: level 5 and level 6, respectively, following the naming convention of the Aquarius project (lower numbers indicate higher resolution). Table 2 gives the dark matter and (initial) gas particle masses for the two resolutions. The gravitational softening is kept fixed in comoving coordinates until redshift  $z_{\text{fix}}$ , after which its value is fixed in physical coordinates. The simulations vary in their choice of  $z_{\text{fix}}$  and therefore the gravitational softening has slightly different values at  $z = 0$ , listed as  $\epsilon_g^{z=0}$  in Table 2 (see also Fig. B1).

## 2.5 Analysis

We describe here some of the conventions and definitions used in the analysis of the simulated galaxies. The centre of the galaxy

is defined to coincide with the position of the baryonic particle with minimum gravitational potential. The *virial radius*,  $r_{200}$ , is the radius of a sphere, centred on the galaxy, with mean density equal to  $200 \rho_{\text{crit}}$ , where  $\rho_{\text{crit}} = 3H^2/8\pi G$  is the critical density for closure. We use the term *halo* to refer to all the mass within  $r_{200}$  and *galaxy* to the baryonic component within a radius  $r_{\text{gal}} = 0.1 r_{200}$  from the centre.

Where a distinction is drawn between ‘hot’ and ‘cold’ gas, we adopt a temperature threshold of  $10^5$  K to separate the two phases. Some codes (e.g. G3, G3-BH, G3-CR, G3-GIMIC, G3-TO and AREPO) adopt an ‘effective’ equation of state to describe the ISM and to circumvent numerical instabilities in poorly resolved regions. This may cause some fluid elements to have nominal temperatures in excess of  $10^5$  K, but still be star forming. In order to prevent assigning this gas to the hot phase, we automatically assign all star-forming gas particles to the ‘cold’ phase.

As we shall see below, different runs yield simulated galaxies of widely varying baryonic mass and angular momentum. In particular, not only the specific angular momentum changes between simulations (as expected, given the wide range in galaxy mass spanned by the various runs), but also its orientation, as we discuss in Appendix C. Because of this, for orientation-dependent diagnostics, we rotate each simulated galaxy to a new coordinate system where the angular momentum vector of its stellar component coincides with the  $z$  direction.

### 3 RESULTS

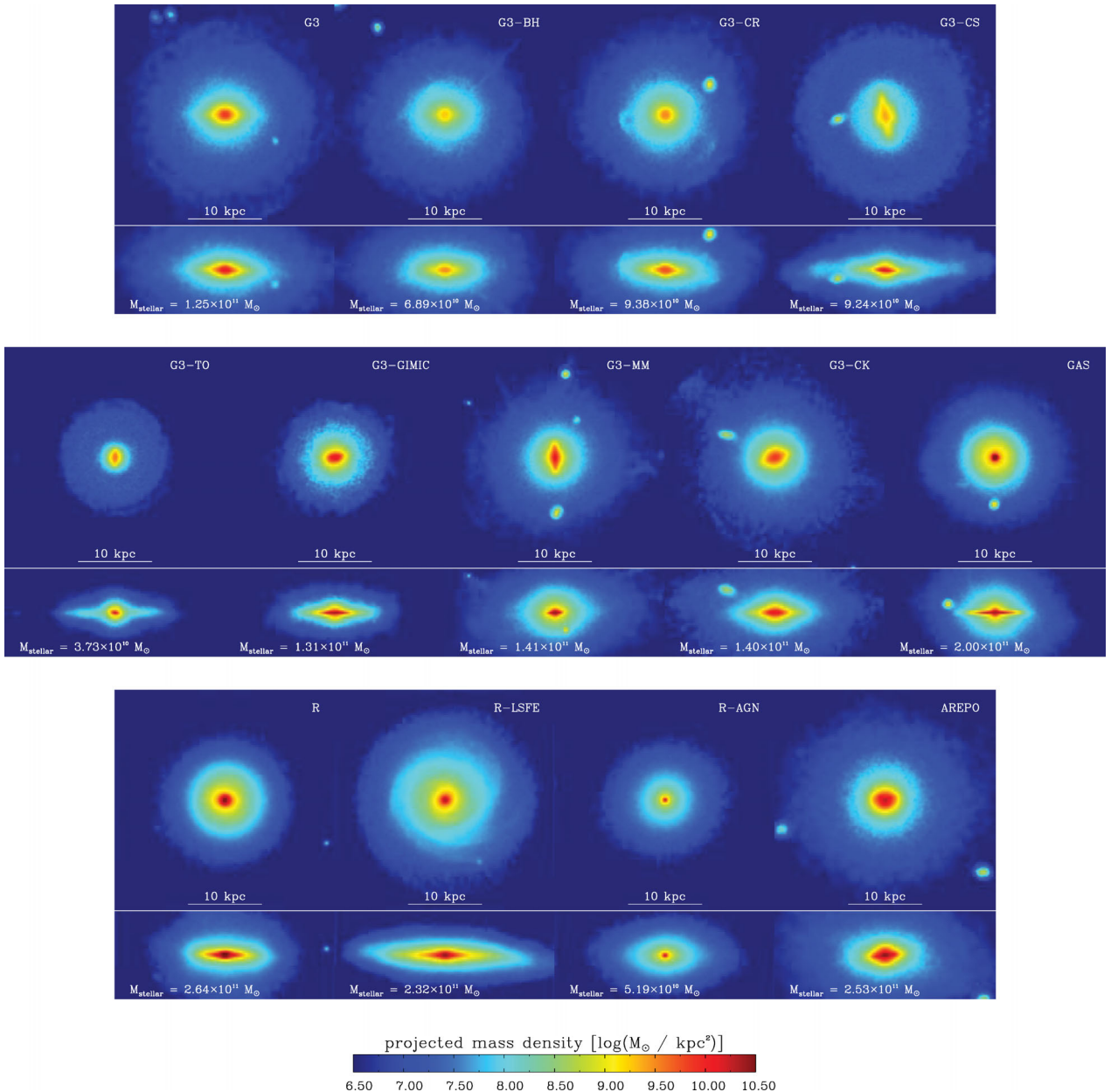
We present here results concerning the stellar mass, morphology, size, star formation history and angular momentum content of the

simulated galaxy. Unless otherwise specified, all results correspond to  $z = 0$  and to the level-5 resolution runs. Numerical convergence between level-5 and level-6 runs is discussed in Section 3.8.

#### 3.1 Galaxy morphology

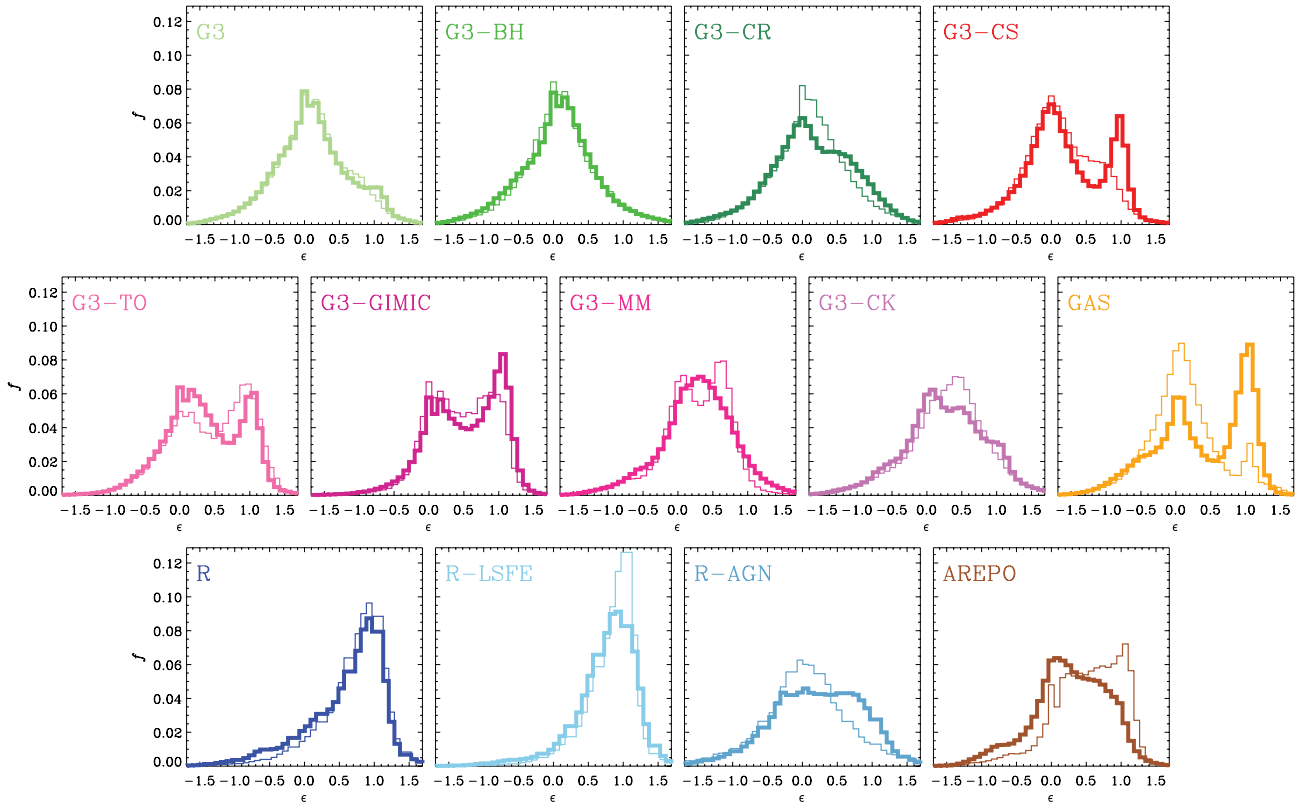
Fig. 2 shows face-on and edge-on maps of the projected stellar mass density for the 13 runs. Labels in each panel list the simulation name, as given in Table 1, as well as the total stellar mass of the galaxy (i.e. within  $r_{\text{gal}}$ ).

These figures illustrate the complex morphology of the simulated galaxies; bars, bulges and extended discs are present, but their relative prominence varies widely from run to run. The galaxy stellar



**Figure 2.** Face-on and edge-on maps of projected stellar mass density. The face-on projection is along the direction of the angular momentum vector of galaxy stars. The face-on and edge-on maps are  $30 \times 30$  and  $30 \times 12 \text{ kpc}^2$ , respectively. The size of each pixel is  $58.6 \text{ pc}$  on a side and its colour is drawn from a logarithmic colour map of the surface stellar mass density. The total stellar mass within the galaxy radius ( $r_{\text{gal}} = 0.1 r_{200} \sim 25 \text{ kpc}$ ) is listed in the legend of each panel.





**Figure 3.** Distribution of stellar circularities,  $\epsilon = j_z/j_c$ , for the different models. The circularity parameter is the  $z$ -component of the specific angular momentum of a star particle,  $j_z$ , expressed in units of the circular orbit value,  $j_c$ , at that radius. Stars with  $\epsilon \approx 1$  typically belong to a rotationally supported disc component. Thick and thin lines correspond to level-5 and level-6 resolution runs, respectively.

mass also shows large scatter, spanning about a decade from the least (G3-TO) to the most massive (R), respectively.

A quantitative measure of the importance of a rotationally supported component is provided by the distribution of stellar *circularities*,  $\epsilon$ , defined as the ratio between the  $z$ -component of the specific angular momentum of a star and that of a circular orbit at the same radius  $r$ :

$$\epsilon = \frac{j_z}{j_c(r)} = \frac{j_z}{r V_c(r)}, \quad (1)$$

where  $V_c(r) = \sqrt{GM(<r)/r}$  is the circular velocity. Stars belonging to a disc are expected to have  $\epsilon \sim 1$ , whereas stars belonging to a non-rotating spheroidal component should have an  $\epsilon$ -distribution roughly symmetric around zero (see e.g. Abadi et al. 2003b; Scannapieco et al. 2009).

We show the circularity distribution of all 13 runs in Fig. 3. Thick and thin lines correspond to the level-5 and level-6 resolution simulations, respectively. The diversity in morphology seen in Fig. 2 is clearly reflected in the distribution of circularities. Thin discs that appear prominently in the images show up as well-defined peaks in the circularity distribution at  $\epsilon \sim 1$ , a distinction that sharpens at higher numerical resolution. In some cases, notably G3, G3-MM, G3-CK and AREPO, the galaxy is noticeably flattened and clearly rotating, but lacks a prominent thin disc.

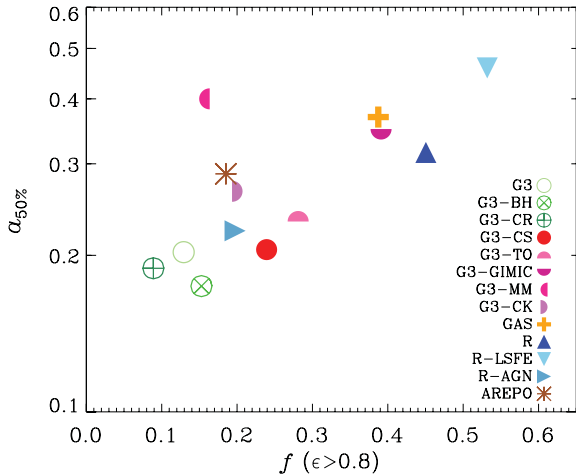
The importance of a thin disc may be crudely estimated by the fraction of stars with  $\epsilon > 0.8$ ,  $f(\epsilon > 0.8)$ .<sup>3</sup> Only in four simu-

lated galaxies do more than  $\sim 40$  per cent of stars satisfy this condition, two SPH based and two AMR based: R, R-LSFE, G3-GIMIC and GAS. The most extreme case, R-LSFE, provides a clue to this behaviour. In this simulation feedback is inefficient and star formation is deliberately delayed, allowing gas to accrete into the galaxy and settle into a centrifugally supported structure before turning into stars.

Indeed, any mechanism that hinders the early transformation of gas into stars without curtailing gas accretion later on is expected to promote the formation of a disc (see e.g. Navarro & Steinmetz 1997). As a result, the galaxies with most prominent discs are also the ones with the youngest stars (Agertz, Teyssier & Moore 2011). This is shown in Fig. 4, where we plot  $f(\epsilon > 0.8)$  versus the median formation time of all stars in the galaxy (expressed in terms of the expansion factor,  $a_{50 \text{ per cent}}$ ). A clear correlation emerges, with discs increasing in prevalence in galaxies that make their stars later. On the other hand, galaxies that make their stars early tend to be spheroid dominated.

An interesting outlier to this trend is G3-MM, which forms stars as late as R but has a small fraction of stars in a disc. Further investigation shows that the G3-MM galaxy did harbour a disc, but it was severely impacted by a collision with a massive satellite in recent times. The satellite is present in other runs, but it has not yet collided with the main galaxy in the majority of cases. This is due to the fact that even small differences in the early evolution get amplified with time and can lead to large discrepancies in the orbital phase of satellites later on. To the extent that this can influence the morphology of the central galaxy, a certain degree of stochasticity in the morphological evolution of a simulated galaxy seems unavoidable.

<sup>3</sup> Note, however, that these fractions often compare poorly with photometric estimates of the disc-to-total ratios (Abadi et al. 2003a; Scannapieco et al. 2010).



**Figure 4.** Median formation time of stars in the galaxy at  $z = 0$ , expressed in terms of the expansion factor,  $a_{50 \text{ per cent}} = 1/(1 + z_{50 \text{ per cent}})$ , as a function of the fraction of stars with circularity exceeding 0.8.

Another interesting result to note is that neither G3 nor AREPO form prominent thin discs. These two runs share the same subgrid physics, but use very different numerical hydrodynamical techniques, which suggests that the morphology of the simulated galaxies is indeed rooted mainly in how gas gets accreted and transformed into stars and in the merger history of the particular halo. As discussed recently by e.g. Torrey et al. (2011), the numerical scheme *does* make a difference when considering the detailed properties of simulated discs, but it does not seem to be the main reason why the G3 and AREPO runs lack discs in this halo. Rather, the failure of feedback to regulate effectively the onset of early star formation and to allow for late gas accretion seems the most likely culprit. Support for this interpretation is provided by G3-BH and G3-CR which, despite the increased feedback, fail to prevent most stars from forming quite early. As a result, none of these models allows a sizeable thin disc to develop.

Models with more efficient feedback schemes, such as those where feedback regulates more effectively early star formation through galactic winds (e.g. G3-CS, G3-TO, G3-GIMIC and GAS), yield galaxies with two well-defined components: an old, non-rotating spheroid surrounded by a young rotationally supported disc. Still, even in this case the disc component is subdominant in terms of total stellar mass, with  $f(\epsilon > 0.8)$  around  $\sim 30\text{--}40$  per cent.

Finally, it is worth noting that the morphology of a galaxy is often dissimilar at the two resolutions attempted here. In general, more prominent discs form at higher resolution but in some cases this trend is reversed. We further discuss resolution effects in Section 3.8.

### 3.2 Rotation curves

As discussed above, regulating star formation without hindering the formation of a stellar disc is a challenging task for any feedback implementation. One might argue that a solution might simply be to delay star formation, such as in R-LSFE, but this comes at the expense of unrealistic disc properties. A simple and convincing diagnostic is the ‘rotation curve’ of the disc which, for simplicity, we represent by the circular velocity profile of the galaxy,  $V_c(r)$ .

This is shown in Fig. 5, where we group in four panels the results of the 13 level-5 Aquila runs, and compare them with the circular velocity curve of the dark-matter-only Aq-C run (dark solid line)

and, for reference, with the rotation curve of the Milky Way as compiled by Sofue et al. (2009).<sup>4</sup>

This figure makes clear that the ‘best discs’ in terms of morphology (i.e. R-LSFE, R, GAS and G3-GIMIC) all have steeply *declining* rotation curves, at odds with the flat rotation curves characteristic of normal spirals. The extreme R-LSFE model again illustrates the problem: here feedback is inefficient at removing baryons, allowing large amounts of gas to collect in a central disc before being turned into stars. A large fraction of these baryons have relatively low angular momentum, however, leading to the formation of a disc that is unrealistically concentrated and with a declining rotation curve. (A similar consideration applies to G3 and AREPO.) It seems one could argue that a successful feedback mechanism must selectively *remove* low angular momentum material from the galaxy (see e.g. Navarro & Steinmetz 1997; van den Bosch et al. 2002; Brook et al. 2011).

Support for this view comes from inspection of the rotation curves of galaxies where galactic winds play a more substantive role, especially at high redshift, when low angular momentum baryons are preferentially accreted: G3-CS and G3-TO show nearly flat rotation curves. A similar result is found for G3-BH, G3-CR and R-AGN, but in these cases the ‘success’ must be qualified by noting that none of these galaxies has a clearly defined stellar disc: the flat  $V_c$  curves are just a reflection of the low baryonic mass of the galaxy that results from adopting these extremely effective feedback models.

### 3.3 Stellar mass

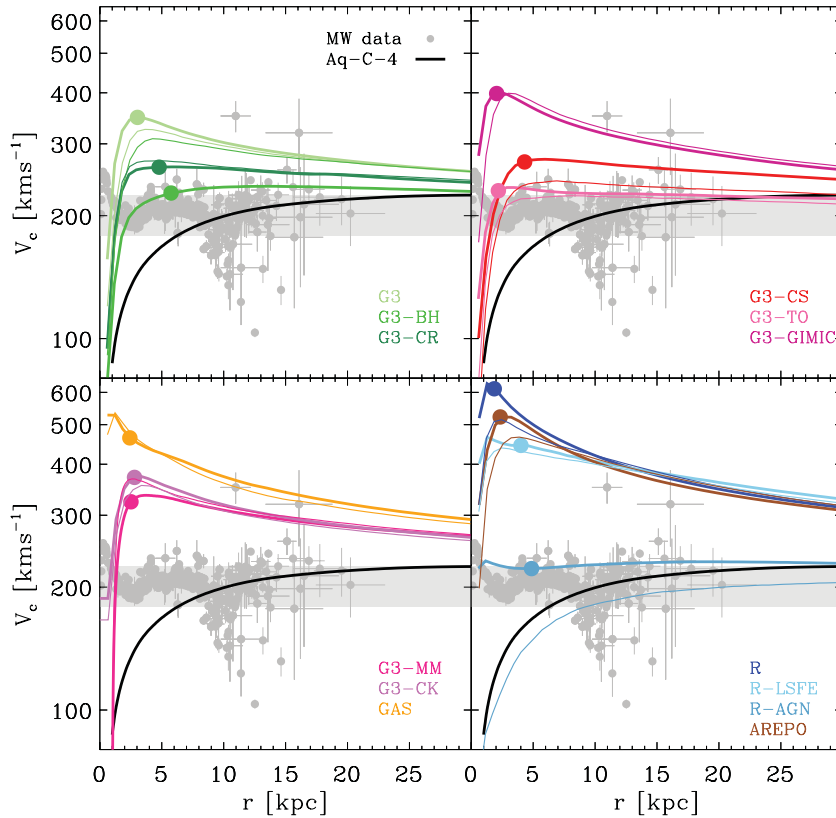
The stellar mass of a galaxy is determined by the combined effects of radiative cooling, the rate at which cold gas is transformed into stars and the ability of feedback to regulate the supply of star-forming gas. Fig. 6 shows how the various implementations affect the stellar mass of the central galaxy,  $M_{\text{stellar}}$ . This figure shows  $M_{\text{stellar}}$  as a function of  $M_{200}(z)$ , the virial mass of the main progenitor from  $z = 2$  to 0. (The symbols correspond to values at  $z = 0$ .)

To guide the interpretation, we show with a dashed curve the total baryonic mass within the virial radius corresponding to the universal baryon fraction,  $(\Omega_b/\Omega_m)M_{200}$ , which sets a hard upper limit to the stellar mass of the central galaxy. The shaded region surrounding the dotted curve corresponds to the stellar masses predicted, at  $z = 0$ , by requiring agreement between the halo mass and galaxy stellar mass functions through simple abundance matching (Guo et al. 2010). We also show the prediction for Aq-C of the semi-analytic models GALFORM (Cooper et al. 2010) and L-GALAXIES (Guo et al. 2011).

The most striking feature of Fig. 6 is the large code-to-code scatter in the stellar mass of the galaxy, which varies between  $\sim 4 \times 10^{10}$  and  $\sim 3 \times 10^{11} M_\odot$  at  $z = 0$  (Table 3). The three largest stellar masses are obtained with the mesh-based codes, R, R-LSFE and AREPO, and correspond to assembling nearly all available baryons in the central galaxy. This illustrates the weak efficiency of the feedback implementations chosen for these codes, aided by the fact that, at comparable resolution, cooling efficiency is enhanced in mesh-based codes relative to SPH (Keres et al. 2011; Sijacki et al. 2011; Vogelsberger et al. 2011).

Indeed, G3 forms only about half as many stars as AREPO, despite sharing exactly the same subgrid physics. The galaxy formed by GAS also has a large stellar mass, but this may be due to the fact that this code has a more efficient cooling function through the addition

<sup>4</sup> Note that Sofue et al. (2009) assume a Galactocentric position and velocity of the Sun of 8 kpc and  $200 \text{ km s}^{-1}$ , respectively.



**Figure 5.** Circular velocity curves of all galaxies in the level-5 runs. The four panels group the results according to numerical technique. The top left-hand panel corresponds to various feedback choices of the standard GADGET code; the top right-hand and bottom left-hand panels correspond to other, independent star formation/feedback modules developed for GADGET, as well as the SPH-based GASOLINE code. The bottom right-hand panel groups the results of the AMR code RAMSES and the moving mesh code AREPO. Thick and thin lines correspond to level-5 and level-6 resolution runs, respectively. The solid circles indicate, for the level-5 simulations, the position of the stellar half-mass radius of each simulated galaxy. The thick black line shows the circular velocity of the dark-matter-only simulation of the same halo (Aq-C). For reference, the region shaded in light grey is bounded by the peak and virial velocities of the Aquarius halo. Dark grey points with error bars are observed data for the Milky Way’s rotation curve, as compiled by Sofue, Honma & Omodaka (2009).

of metal-line cooling. Although the numerical technique may affect some changes in the stellar mass, these are small compared with the variations introduced by the feedback implementation. This may be seen by noting that, when including active galactic nucleus (AGN) feedback in RAMSES, the stellar mass decreases by a factor of  $\sim 5$  and the disc component is largely erased.

Note that the large variations in the stellar mass of the galaxy formed in different runs imply that the dark halo will respond by contracting differently in each case, as we discuss in Section 3.4. Note also that the differences in stellar mass are dominated by differences prior to  $z = 2$ ; in fact, in some simulations the stellar mass at  $z = 2$  is already above the  $z = 0$  stellar mass–halo mass relation.

It is also important to note that feedback *must* be roughly as effective as that of R-AGN in order to obtain stellar masses consistent (within the error) with the abundance-matching predictions. Indeed, the only other codes to match this constraint, and thus fall within the shaded area of Fig. 6 are G3-BH and G3-TO; of these only the latter forms a galaxy with a discernible disc (see Fig. 2). All other models give stellar masses well in excess of the abundance-matching constraint, a shortcoming of most published galaxy formation simulations to date (Guo et al. 2010; Sawala et al. 2011).

It is also worth noting that the abundance-matching models allow for substantial scatter in the  $M_{200}-M_{\text{stellar}}$  relation. Indeed, the

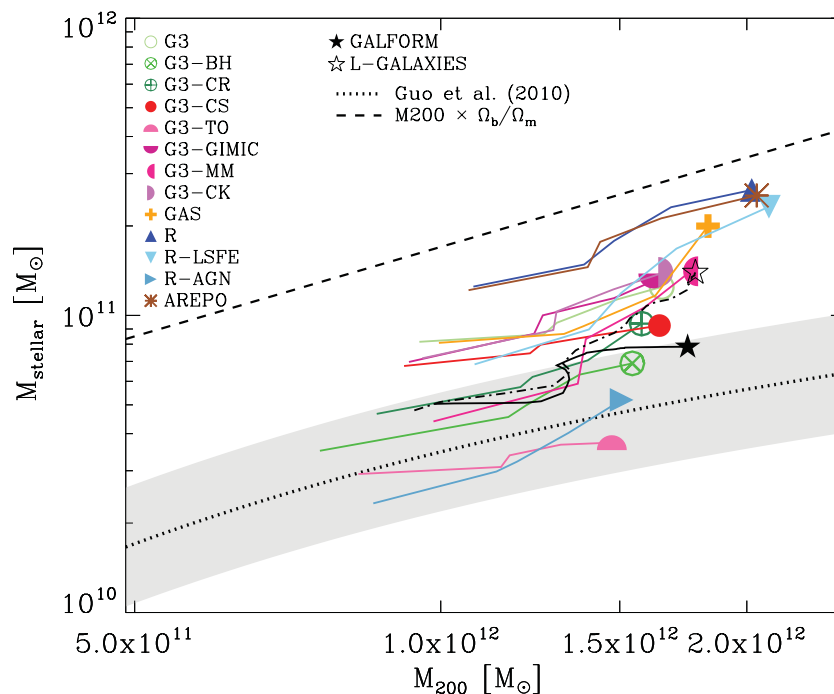
more sophisticated treatments of the L-GALAXIES and GALFORM semi-analytic codes indicate that Aq-C might form a galaxy more massive than expected on average for a halo of that mass (see open and filled starred symbols in Fig. 6). L-GALAXIES, in particular, suggests that Aq-C might be a  $2\sigma$  outlier from the relation, which would alleviate, but not resolve, the disagreement between the results of R, R-LSFE, AREPO and GAS and the model predictions. GALFORM, on the other hand, predicts that Aq-C should be about  $1\sigma$  above the mean abundance-matching relation.

Taken altogether, these results illustrate the basic challenge faced by disc galaxy formation models: feedback must be efficient enough either to prevent the accretion, or to facilitate the removal, of most baryons, whilst at the same time allowing enough high angular material to accrete and form an extended stellar disc.

### 3.4 Tully–Fisher relation

The stellar mass and circular velocity of disc galaxies are strongly linked by the Tully–Fisher relation, and it is therefore instructive to compare the properties of simulated galaxies with those of observed discs. This is done in Fig. 7, where we compare data compiled by Dutton et al. (2011) from Pizagno et al. (2007), Verheijen (2001) and Courteau et al. (2007) with the 13 simulated galaxies.





**Figure 6.** The stellar mass of the central galaxy as a function of the virial mass of the surrounding halo. Curves of different colour track the evolution of the galaxy in each simulation between  $z = 2$  and 0. The dotted line indicates the stellar mass *expected* at  $z = 0$  from the abundance-matching analysis of Guo et al. (2010); the shaded region corresponds to a 0.2 dex uncertainty. The dashed line indicates the mass of all baryons within the virial radius,  $(\Omega_b/\Omega_m) M_{200}$ . The filled and open star symbols indicate the predictions of the semi-analytic models GALFORM (Cooper et al. 2010) and L-GALAXIES (Guo et al. 2011) for halo Aq-C, respectively. The dot-dashed curves show the evolution since  $z = 2$  according to these two models.

Because the rotation curves of simulated galaxies are not flat (see Fig. 5), we use velocities estimated at the stellar half-mass radius in order to be as consistent as possible with the rotation speeds estimated observationally from spatially resolved rotation curves (see e.g. Courteau et al. 2007). The symbols connected by a solid line show the contribution of the dark matter to the circular velocity *at the same radius*. A dotted line shows the same, but for the dark-matter-only Aq-C halo; the difference between solid and dotted curves indicates the degree of ‘contraction’ of the dark halo.

There is a clear discrepancy between the observed Tully–Fisher relation and simulated galaxies, which tend to have substantially larger velocities at given  $M_{\text{stellar}}$ . The disagreement worsens for large stellar masses, emphasizing again the fact that too many baryons are able to cool and form stars in these systems. Interestingly, at low stellar mass simulated galaxies approach the observed relation but still have, on average, higher rotation speeds than typical discs. This suggests that, although these galaxies may have stellar masses consistent with abundance-matching considerations (see Section 3.3), they must differ from typical spirals in other respects, such as an excessive concentration of the dark matter or luminous component.

The dark matter *contribution* to the circular velocity (connected symbols in Fig. 7) lies well below the average rotation speed expected from the Tully–Fisher relation. This suggests that the concentration of dark matter is not the origin of the disagreement; there should in principle be no problem matching the observed relation provided that the luminous component of the galaxy is extended enough. The offset from the observed Tully–Fisher relation thus suggests that simulated galaxies are more concentrated than normal spirals, resulting in discs that rotate too fast for their stellar mass. We analyse the size of simulated galaxies in Section 3.6, after examining the importance of the gaseous component of the galaxy next.

### 3.5 Gaseous component

Fig. 8 shows  $f_{\text{gas}}$ , the fraction of the baryonic mass of simulated galaxies in form of gas at  $z = 0$ , as a function of the  $R$ -band absolute magnitude, and compares them with data for star-forming galaxies from Schombert, McGaugh & Eder (2001), Bell & de Jong (2000) and Haynes et al. (1999). Magnitudes for the simulations were calculated using the dust-free Bruzual & Charlot (2003) population synthesis models, for a Chabrier initial mass function (IMF) and solar metallicity.

Most simulated galaxies have gas fractions below 10 per cent, which puts them at odds with observations of nearby spirals. As for the stellar mass, note the large code-to-code scatter in  $f_{\text{gas}}$ , which varies from about 1 per cent for G3 to nearly 30 per cent for R-LSFE. As expected, galaxies with larger gas fractions are predominantly those with morphologies that include a well-defined stellar disc, presumably because of ongoing star formation. The converse, however, is not always true: G3-CS and R have low gas fractions at  $z = 0$  but prominent discs.

More surprisingly perhaps, the gas fraction seems to correlate only weakly with the present-day SFR (which we discuss more thoroughly in Section 3.7). For example, R-AGN has the third largest gas fraction and by far the lowest SFR at  $z = 0$ . The same applies to G3-TO, which, despite its large  $f_{\text{gas}}$ , forms stars at rates well below what would be expected for an average spiral (see Fig. 12).

Overall we see no obvious dependence of the gas fraction on the numerical method: of the four galaxies with highest  $f_{\text{gas}}$ , two are SPH based (G3-TO and G3-MM) and two are AMR based (R-AGN and R-LSFE). However, we note that AREPO has a much higher gas fraction (and stellar mass) than G3, despite sharing the same subgrid physics. This supports the conclusion that standard SPH-based methods may underestimate the total amount of gas that

**Table 3.** Properties of simulated galaxies at  $z = 0$ , for the level-5 (upper row) and level-6 (lower row) resolution simulations. We also list the prediction of the semi-analytic models GALFORM (Cooper et al. 2010) and L-GALAXIES (Guo et al. 2011), and of the dark-matter-only simulation Aq-C-4 of the Aquarius Project.

Code	$M_{200}$ ( $10^{10} M_{\odot}$ )	$r_{200}$ (kpc)	$V_{200}$ ( $\text{km s}^{-1}$ )	$M_{\text{stellar}}$ ( $10^{10} M_{\odot}$ )	$M_{\text{cold gas}}$ ( $10^{10} M_{\odot}$ )	$f(\epsilon > 0.8)$	$a_{50 \text{ per cent}}$	$R_{\text{h, stars}}$ (kpc)	$R_{\text{h, gas}}$ (kpc)	$V_{1/2}$ ( $\text{km s}^{-1}$ )	SFR ( $M_{\odot} \text{ yr}^{-1}$ )	$M_{\text{R}}$	$f_{\text{gas}}$	$V_{\text{DM}(R_{\text{h}})}/V_{\text{C}(R_{\text{h}})}$
G3	164.94	239.12	172.24	12.47	0.063	0.13	0.20	3.06	3.06	348.19	0.378	-21.25	0.011	0.521
	172.39	242.23	174.95	11.54	0.003	0.11	0.20	3.43	1.00	325.44	0.906	-21.17	0.046	0.552
G3-BH	157.61	233.48	170.39	6.89	0.161	0.15	0.17	5.31	7.42	224.88	0.333	-20.51	0.026	0.695
	167.38	242.35	172.34	11.07	0.027	0.10	0.19	3.43	8.36	304.63	0.242	-20.99	0.070	0.557
G3-CR	154.32	234.29	168.31	9.38	0.247	0.09	0.19	4.75	7.41	262.93	0.333	-20.93	0.029	0.639
	172.85	239.25	176.27	9.22	0.001	0.09	0.19	4.28	2.45	272.73	0.000	-20.77	0.016	0.652
G3-CS	164.11	237.45	172.41	9.24	0.200	0.24	0.21	4.27	18.22	270.82	0.341	-20.85	0.024	0.619
	149.46	230.01	167.17	5.28	0.061	0.13	0.18	3.79	7.37	236.04	0.065	-20.19	0.018	0.706
G3-TO	147.32	228.40	166.56	3.73	0.850	0.28	0.24	1.94	14.30	213.98	0.253	-20.32	0.187	0.533
	147.85	228.70	166.74	3.94	0.084	0.35	0.24	3.39	3.39	216.65	0.199	-20.43	0.028	0.677
G3-GIMIC	161.84	235.76	171.82	13.06	0.982	0.39	0.34	1.95	10.39	398.37	3.425	-22.08	0.072	0.353
	167.57	238.41	173.86	14.10	1.082	0.31	0.34	2.44	5.97	388.78	4.635	-22.18	0.073	0.355
G3-MM	176.11	245.02	175.82	14.07	0.205	0.16	0.31	3.96	3.96	335.49	5.471	-22.43	0.078	0.542
	176.69	242.57	177.00	13.74	0.318	0.11	0.34	2.28	2.54	362.90	6.138	-22.33	0.087	0.374
G3-CK	166.45	237.61	173.57	14.00	0.324	0.20	0.27	3.52	8.46	373.97	3.637	-22.12	0.025	0.444
	177.52	243.66	177.01	12.30	0.263	0.18	0.26	3.18	3.96	346.24	3.633	-22.15	0.033	0.421
GAS	183.18	246.12	178.91	19.98	0.749	0.39	0.37	3.55	5.52	440.25	13.892	-23.17	0.101	0.427
	183.31	246.23	178.93	17.31	1.090	0.12	0.33	2.85	3.55	505.07	18.574	-22.90	0.147	0.385
R	202.25	256.20	184.26	26.43	1.084	0.45	0.32	2.56	4.48	580.44	5.327	-22.61	0.022	0.387
	178.74	244.00	177.49	26.67	2.063	0.51	0.35	3.66	10.37	504.81	7.650	-22.66	0.054	0.457
R-LSFE	210.27	258.70	186.97	23.21	3.101	0.53	0.46	4.53	9.06	444.55	12.704	-22.96	0.267	0.464
	180.34	245.10	177.89	23.28	3.450	0.62	0.44	5.51	12.25	428.59	6.284	-22.73	0.199	0.554
R-AGN	150.63	231.80	167.17	5.19	0.512	0.20	0.22	5.22	16.23	222.24	0.068	-20.28	0.128	0.677
	147.61	229.10	166.46	1.50	0.319	0.11	0.21	6.87	14.89	169.58	0.000	-18.86	0.014	0.834
AREPO	204.54	257.45	184.85	25.33	0.382	0.19	0.29	2.21	5.47	498.77	8.827	-22.59	0.040	0.343
	206.21	257.54	185.57	28.68	0.951	0.36	0.37	3.48	8.61	464.29	11.364	-22.88	0.051	0.416
GALFORM	203.27	261.01	183.01	7.84	0.004		0.26	3.77	10.43		0.004	-20.99	0.001	
L-GALAXIES	178.01	243.10	177.46	13.95	2.44		0.44	2.03	5.13		10.328	-22.80	0.149	
AQ-C-4	179.30	243.68	177.92											

cools and becomes available for star formation, especially when feedback is as weak as implemented in the G3 and AREPO runs (see also Agertz et al. 2007; Vogelsberger et al. 2011).

The interpretation of these results is not straightforward. The gas content of a galaxy is constantly evolving; supplied by accretion, depleted by star formation and removed by feedback-driven winds. This leads to large fluctuations in the instantaneous gas fraction and SFR of a galaxy, which may be exacerbated by chance events such as satellite accretion.

### 3.6 Galaxy size

The left-hand panel of Fig. 9 compares the stellar half-mass radii of simulated galaxies with the size of observed galaxies of similar stellar mass,<sup>5</sup> as well as with the predictions of the GALFORM and L-GALAXIES semi-analytic models and the Milky Way, for reference. The observed sizes correspond to Petrosian half-light radii in the  $r$  band rather than stellar half-mass radii so the comparison should only be taken as indicative. Red/blue dots correspond to galaxies redder/bluer than  $(g - r) = 0.59 + 0.052 \log_{10}(M_{\text{stellar}}/M_{\odot}) - 10.0$

<sup>5</sup> Data taken from the SDSS MPA-JHU Data Release 7 (DR7) for nearby ( $z < 0.1$ ) galaxies; <http://www.mpa-garching.mpg.de/SDSS/DR7/>

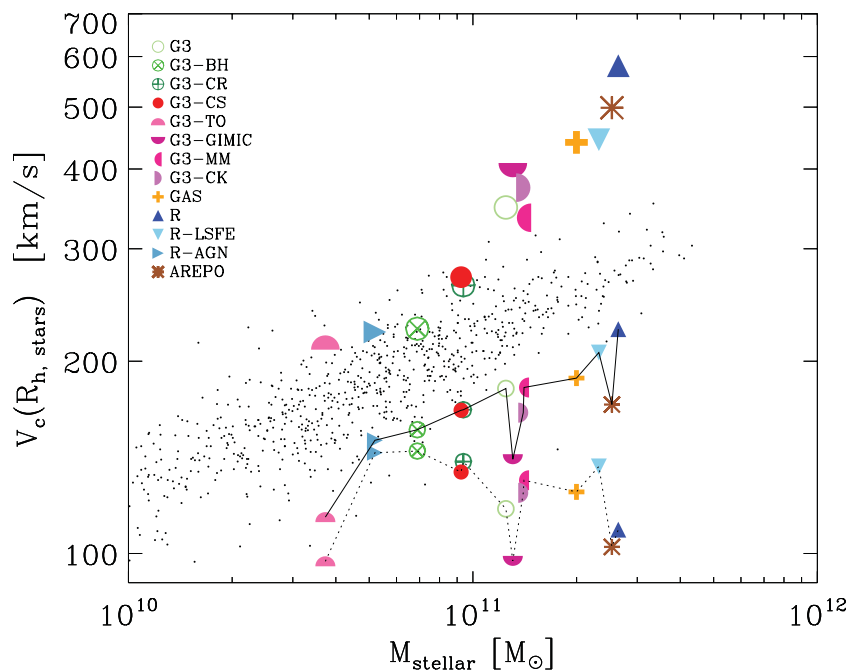
and are meant to outline roughly the location of early- and late-type galaxies in this plot.

As anticipated in the previous subsection, most galaxies are too compact to be consistent with typical spiral galaxies of comparable stellar mass. In general, the more massive the simulated galaxy, the smaller its size, a trend that runs contrary to observation. In particular, the most massive simulated galaxies (R, R-LSFE, AREPO) are even smaller than most early-type galaxies in the nearby Universe, highlighting again the shortcomings of simulations where cooling and star formation proceed largely unimpeded by feedback.

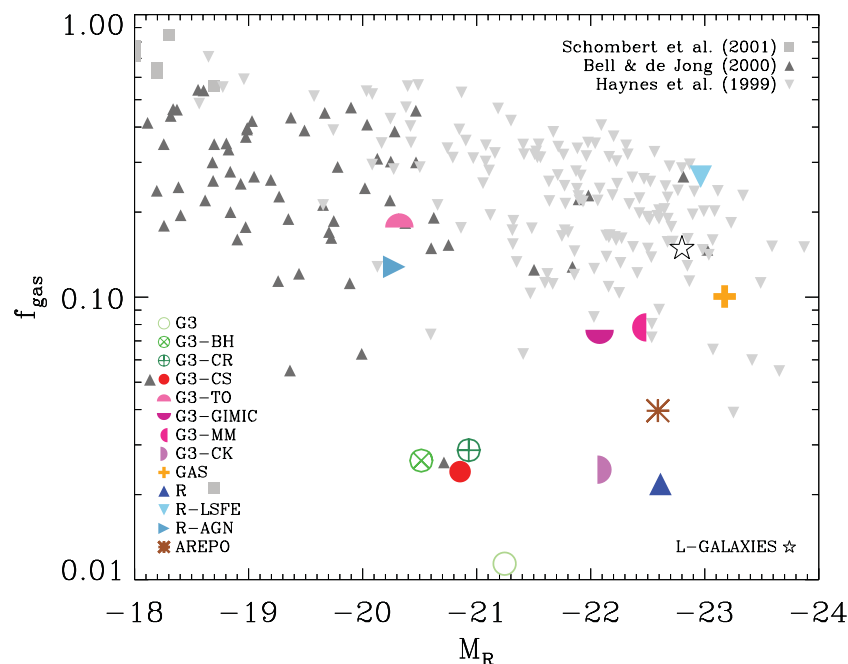
Simulations where feedback is more effective at curtailing the formation of stars give rise to galaxies with sizes in better agreement with observation. For example, the half-mass radii of R-AGN and G3-BH reach  $R_{\text{h}} \sim 5$  kpc for  $\sim 7 \times 10^{10} M_{\odot}$ , at the lower end of the distribution of spiral sizes. However, as discussed in Section 3.1, neither of these galaxies has a disc-like morphology.

The simulated galaxy with lowest  $M_{\text{stellar}}$  and a well-defined disc is G3-TO, but, as seen from Fig. 9, it has a half-mass radius of less than 2 kpc, well below what would be expected for a spiral of that mass. This is because most of the stellar mass in G3-TO is in the form of a highly concentrated spheroid rather than in an extended disc. Therefore, even in this case, feedback has apparently allowed too many low angular momentum baryons into the galaxy.

These results support our earlier conclusion: feedback must not only limit how many baryons settle into the galaxy, but must also



**Figure 7.** The Tully–Fisher relation. The circular velocity at the stellar half-mass radius of each simulated galaxy is plotted as a function of stellar mass for all 13 level-5 runs. Small black dots correspond to data for nearby spirals taken from Pizagno et al. (2007), Verheijen (2001) and Courteau et al. (2007). The symbols connected by a solid line show the contribution of the dark matter to the circular velocity at  $R_{h, \text{stars}}$ . Those connected by the dotted line show the circular velocity of the dark-matter-only halo (Aq-C) at the same radii.

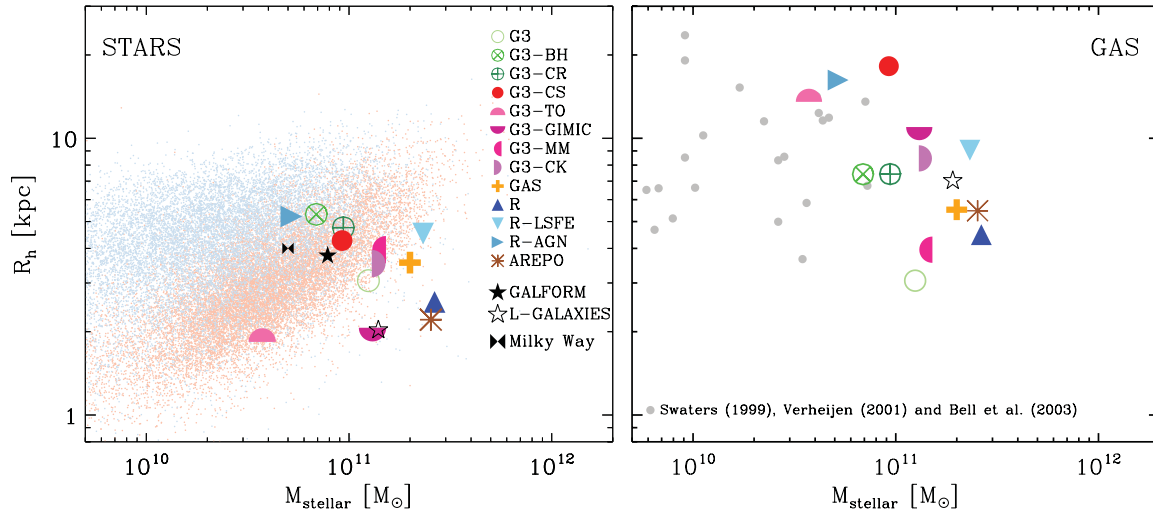


**Figure 8.** Gas mass fraction,  $f_{\text{gas}} = M_{\text{gas}} / (M_{\text{gas}} + M_{\text{stellar}})$ , of the galaxy versus  $R$ -band absolute magnitude. Magnitudes have been calculated using the Bruzual & Charlot (2003) population synthesis models for solar metallicity and a Chabrier IMF and ignoring the effects of dust extinction. Symbols in grey/black show data for nearby spirals compiled from the references listed in the figure label. We also show the cold gas fraction prediction of the semi-analytic model L-GALAXIES (Guo et al. 2011) for Aq-C. The gas fraction predicted by GALFORM (Cooper et al. 2010) is close to zero and thus lies outside the plotted range.

selectively allow high angular momentum material to be accreted and retained in order to form a realistic disc.

The right-hand panel of Fig. 9 shows the projected half-mass radius of cold gas in the simulated galaxies (at  $z = 0$ ) as a function of stellar mass and compares them with H I observations compiled by

Dutton et al. (2011) from Swaters et al. (1999) and Verheijen (2001). Despite the large code-to-code variation, the simulated gaseous discs are systematically more extended than the stellar component, in agreement with observation. They are also in better agreement overall with the typical size of H I discs, a result that suggests



**Figure 9.** Left: stellar half-mass radius as a function of the stellar mass of the galaxy. Blue and red dots show the Petrosian  $r$ -band half-light radius for a sample of nearby ( $z < 0.1$ ) SDSS galaxies taken from the MPA-JHU DR7 release. The sample is split into ‘blue cloud’ and ‘red sequence’ galaxies depending on their colours, according to the condition  $(g - r) = 0.59 + 0.052 \log_{10}(M_{\text{stellar}}/M_{\odot}) - 10.0$  (only 5 per cent of randomly selected data points are shown). We also show the predictions of the semi-analytic models GALFORM (Cooper et al. 2010) and L-GALAXIES (Guo et al. 2011) for Aq-C, and the approximate location of the Milky Way in this plot. Right: the projected half-mass radius of cold ( $T < 10^5$  K) gas in the galaxy as a function of stellar mass. Grey circles indicate the half-mass radii of H I discs compiled by Dutton et al. (2011) from Swaters et al. (1999) and Verheijen (2001), and the open star symbol indicates the prediction of L-GALAXIES. The prediction of GALFORM is not shown here, since it predicts a present-day gas mass close to zero.

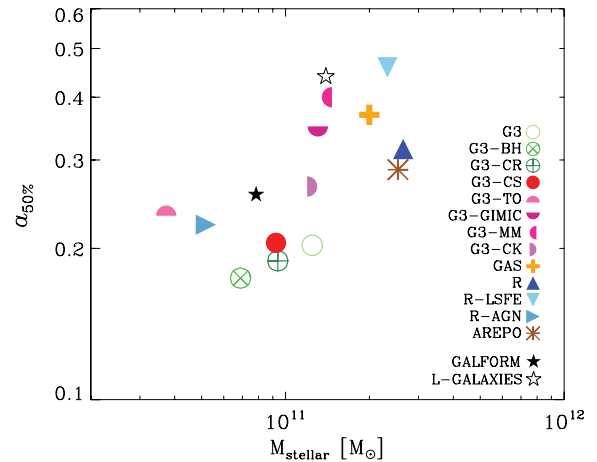
that material accreted relatively recently (and thus still in gaseous form) has, on average, enough angular momentum to form discs of realistic size. If feedback were able to favour the accretion and retention of this late-accreting, high angular momentum gas, then simulated galaxies would have a much better chance of forming realistic discs.

### 3.7 Star formation history

A recurring theme of our discussion so far has been the need to prevent the assembly of an overly massive galaxy without, at the same time, preventing high angular momentum, late-accreting material from reaching the galaxy and forming a disc. This requires feedback to act especially strongly at high redshift, when a large fraction of baryons first become cold and dense enough to start forming stars. None of the Aquila runs seems able to meet these requirements satisfactorily, as shown by the star formation history of the simulated galaxies.

Fig. 10 shows the stellar mass of the galaxy versus the median formation time of the stars (expressed in terms of expansion factor,  $a = 1/(1+z)$ ). Note the strong correlation between the two, which indicates that the codes best able to limit the growth of the mass of the galaxy do so at the expense of curtailing the incorporation of *late-accreting* material. Indeed, the three galaxies with the lowest  $M_{\text{stellar}}$  (G3-TO, R-AGN, G3-BH) form half of their stars in the first Gyr or so of evolution, i.e. by  $z \sim 4$ . It is not surprising then that two of them lack a discernible disc, and that the disc in G3-TO is overwhelmed by a massive, dense spheroid composed mainly of old stars.

Further details on the star formation history of individual galaxies are presented in Fig. 11, where we show, in cumulative and differential form, the distribution of stellar ages of the stars in the galaxy at  $z = 0$ . Note that these are *not* SFRs for the main progenitor, since stars may (and some of them, indeed, do) form in different progenitors before being accreted into the galaxy. Nevertheless, the data in Fig. 11 show clearly that few codes are able to prevent the

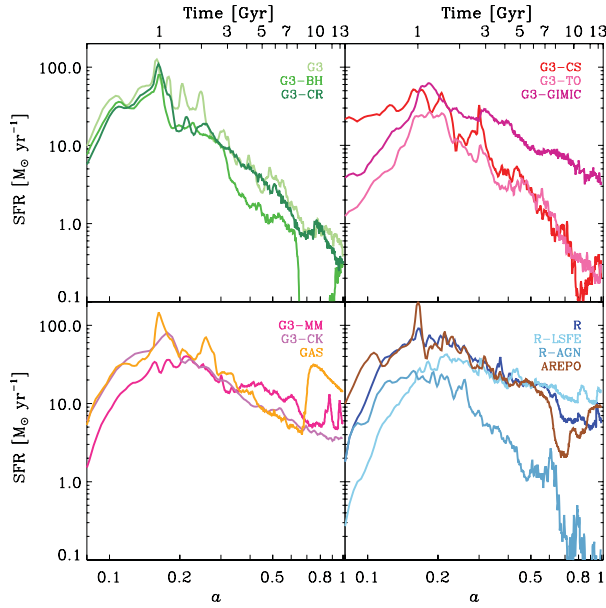


**Figure 10.** Median formation time of stars in the galaxy (expressed in terms of the expansion factor ( $a_{50\text{ percent}}$ )) as a function of total stellar mass at  $z = 0$ . We also show the prediction of the semi-analytic models GALFORM (Cooper et al. 2010) and L-GALAXIES (Guo et al. 2011) for Aq-C.

early burst of star formation activity that accompanies the collapse of the first massive progenitors of the galaxy. Only G3-TO, G3-MM, R-AGN and R-LSFE are successful at keeping this peak ‘rate’ at less than  $\sim 100 M_{\odot} \text{ yr}^{-1}$  at  $z \sim 4-5$ , but even they see a precipitous decline in star formation afterwards. (The exception is R-LSFE, but this is achieved by artificially delaying star formation, see Section 2.1.)

Fig. 11 also shows that the morphological appearance of simulated galaxies is very weakly correlated with star formation history. There are examples of galaxies with lots of recent star formation that have well-defined discs (e.g. R) and examples which do not (e.g. AREPO), as well as cases of galaxies with very little recent star formation that have well-defined discs (e.g. G3-CS, G3-TO) and cases which do not (R, G3-BH).

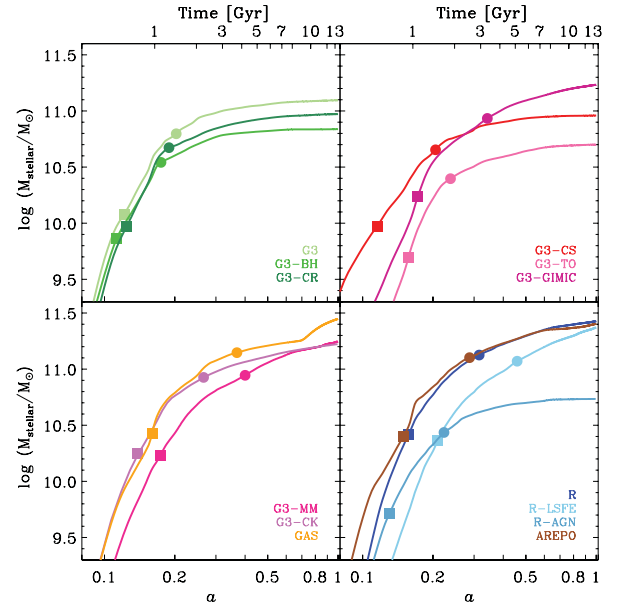
Aside from these general considerations, the details of the star formation history of each galaxy reflect the particular subgrid physics



**Figure 11.** Distribution of stellar formation times (expressed in terms of the expansion factor,  $a$ ). Left- and right-hand panels show the same data, in differential and cumulative form, respectively. The simulations are grouped according to numerical technique, as in Fig. 5. The squares and circles indicate the time of formation of the first 10 and 50 per cent of the stars.

implementation chosen for each code (see Schaye et al. 2010). For example, the feedback scheme of G3-GIMIC (where SN-driven winds are assumed to be launched with fixed speed) imply that it is effective at early times, when the mass of the progenitors is small, but becomes ineffective when the main halo reaches  $\sim 10^{12} M_{\odot}$ . This curtails early star formation but allows the stellar mass to increase substantially at low redshift. On the other hand, feedback in G3-TO is effective at all redshifts, since its strength scales with the potential well of the galaxy. In G3-CS star formation is not effectively regulated at very early times because SN energy feedback is not injected into the ISM instantaneously but rather after a time delay which depends on the local properties of the cold and hot neighbouring particles.

These choices imply that at any given time there is substantial scatter in the star-forming properties of simulated galaxies, compounded by the fact that there is a certain degree of stochasticity in the rate at which a galaxy accretes mass (Section 3.1). For example, in the case of AREPO, the infall of a satellite at  $z \sim 0.7$  disrupts the gas disc and leads to a significant increase in the star-forming activity at  $z = 0$ . In the GAS run, a large burst of star formation also occurs near  $z = 0$ , but in this case it seems associated with enhanced gas accretion facilitated by the infall of a satellite. Such effects are at least partially responsible for the large code-to-code scatter in the SFR of the galaxy at the present time. This is shown in Fig. 12, where we compare the present-day SFR (averaged over the past 0.5 Gyr to smooth out short-term fluctuations) with the stellar mass of simulated galaxies and contrast them with observations of local SDSS galaxies. The observational sample corresponds to nearby ( $z < 0.1$ ) SDSS galaxies selected from the MPA-JHU DR7 catalogue.<sup>6</sup> The SFR of simulated galaxies varies from a low of  $\sim 0.02 M_{\odot} \text{ yr}^{-1}$  (R-AGN) to nearly  $20 M_{\odot} \text{ yr}^{-1}$  (GAS), spanning the whole range covered by observed galaxies, from ‘red and dead’ spheroids to actively star-forming gas-rich discs. The large scatter leads us to conclude that caution must be exercised when analysing the instantaneous



**Figure 12.** Present-day SFR (averaged over the last 0.5 Gyr to smooth out short-term fluctuations) as a function of stellar mass. Blue and red dots correspond to a sample of nearby ( $z < 0.1$ ) SDSS galaxies selected from the MPA-JHU DR7 catalogue (only 5 per cent of randomly selected data points are shown). The sample is split into ‘red sequence’ and ‘blue cloud’ galaxies as described in Fig. 9. We also show the prediction of the semi-analytic model L-GALAXIES of Guo et al. (2011) for Aq-C and the approximate location of the Milky Way (Oliver et al. 2010; Leitner & Kravtsov 2011). Note that the GALFORM semi-analytic model is not shown here since it predicts a present-day SFR close to zero.

SFRs of simulated galaxies, since these depend sensitively on the details of accretion and of the implemented subgrid physics.

### 3.8 Numerical convergence

Convergence is a necessary (but not sufficient) criterion to assess the robustness of numerical simulation results. We discuss here the effects of resolution by comparing the results of level-5 and level-6 simulations for each code (see also Table 3). As a quick reminder, at level 5 the parent halo of the Aquila project, Aq-C, has  $\sim 700\,000$

<sup>6</sup> <http://www.mpa-garching.mpg.de/SDSS/DR7/>



dark matter particles within the virial radius; this number is reduced to  $\sim 90\,000$  at level 6.

Also, when considering convergence one must note that resolution-dependent behaviour is inevitable to some degree, since the Jeans length and Jeans mass of star-forming gas are not well resolved either at level 6 or at level 5. For instance, in several codes stars are allowed to form only above a density threshold of  $n_{\text{th}} \sim 0.1\text{ cm}^{-3}$  and at temperatures of order  $\sim 10^4\text{ K}$ , which correspond to a Jeans length of  $\sim 1.5\text{ kpc}$ , only slightly larger than the gravitational softening at level 6.

Although all codes start from initial conditions of the same mass and spatial resolutions, each code then adopts its own refinement strategy, which may result in significant effective runtime resolutions. For example, some codes of similar type choose different gravitational softening lengths, and differ as well in their choice of when and whether to keep it fixed in comoving or physical coordinates (Fig. B1). This makes it even harder to disentangle numerical resolution effects from code-type effects. The enhanced cooling in grid-based codes discussed in Section 3.3, for example, seems driven by differences inherent to the hydrodynamical treatment, and not by resolution (see Vogelsberger et al. 2011, for details), but the distinction is less clear for other quantities.

We quantify the effects of resolution on a given quantity,  $Q$ , by the fractional variation, measured at  $z = 0$ , between the level-5 and level-6 runs,

$$\Delta Q/Q = (Q_6 - Q_5)/Q_5. \quad (2)$$

Note that with this definition a quantity that increases with increasing resolution will have a *negative* value of  $\Delta Q/Q$ .

We summarize the results in Fig. 13, where we show the variations: (i) in global *galaxy* properties such as peak circular velocity,  $V_{\text{max}}$ , and the radius at which it is achieved,  $r_{\text{max}}$ ; (ii) in properties of the *stellar component*, such as the total mass,  $M_{\text{stellar}}$ , and the half-mass radius,  $R_{\text{h,stars}}$ ; (iii) in properties of the *gas component*, such as the mass,  $M_{\text{gas}}$ , and half-mass radius,  $R_{\text{h,gas}}$ ; (iv) in properties related to *star formation*, such as the gas fraction and the present-day SFR and (v) in the details of the *galaxy assembly/morphology*, such as the median formation time of stars at  $z = 0$  (expressed in terms of the expansion factor,  $a_{50\text{ per cent}}$ ) and the fraction of stars kinematically associated with a rotationally supported disc ( $f(\epsilon > 0.8)$ , see Section 3.1). When differences exceed 100 per cent (i.e.  $|\Delta Q/Q| > 1$ ), we use arrows to indicate if such deviations occur along the  $x$  and/or  $y$  coordinates of the plot. The actual values of all quantities plotted in Fig. 13 are listed in Table 3.

Aside from the fact that numerical convergence is not particularly good for any of the codes, Fig. 13 illustrates a few interesting points. The first concerns the properties of the stellar component. In particular, the total mass in stars and their median age seem to be the most reliable results, suggesting that the star formation/feedback scheme chosen for each code is reasonably independent of resolution. Most codes reproduce the total stellar mass and  $a_{50\text{ per cent}}$  to better than 20–30 per cent. Interestingly, the peak circular velocity is one of the properties least affected by resolution (see also Fig. 5). This is encouraging, since it implies that diagnostics such as the observed Tully–Fisher relation may be used to assess the success of a particular model.

Convergence deteriorates when considering the detailed properties of the stellar component, such as the fraction of stars in high-circularity orbits:  $f(\epsilon > 0.8)$ . Although half the codes give results that converge to better than  $\sim 30$  per cent the variations can be much larger for some codes. Interestingly, *increasing the resolution does not always leads to better-defined discs*: for example, the

fraction of ‘disc’ stars increases by a large fraction in the case of GAS but actually decreases for AREPO. Finally, there is some indication that the most extreme feedback models (i.e. those that result in the lowest  $M_{\text{stellar}}$ ; e.g. R-AGN and G3-BH) are the most vulnerable to resolution-induced changes.

Even larger variations are seen for the properties of the gas component: only five of the 13 simulations show variations in  $M_{\text{gas}}$  and  $R_{\text{h,gas}}$  smaller than 50 per cent, and three simulations have differences in  $M_{\text{gas}}$  larger than 100 per cent. Similar results are found for the gas fractions, and, consequently, for the present-day average SFRs. In general, increasing the resolution leads to larger gaseous discs, but in some cases the total mass in gas increases while it decreases in others. These large variations are at least partly due to the fact that some galaxies (e.g. R-AGN) have almost no gas left at  $z = 0$ , and therefore even small changes can lead to large fractional variations. The properties of the gaseous component seem the most vulnerable to numerical resolution effects and therefore caution must be exercised in their interpretation.

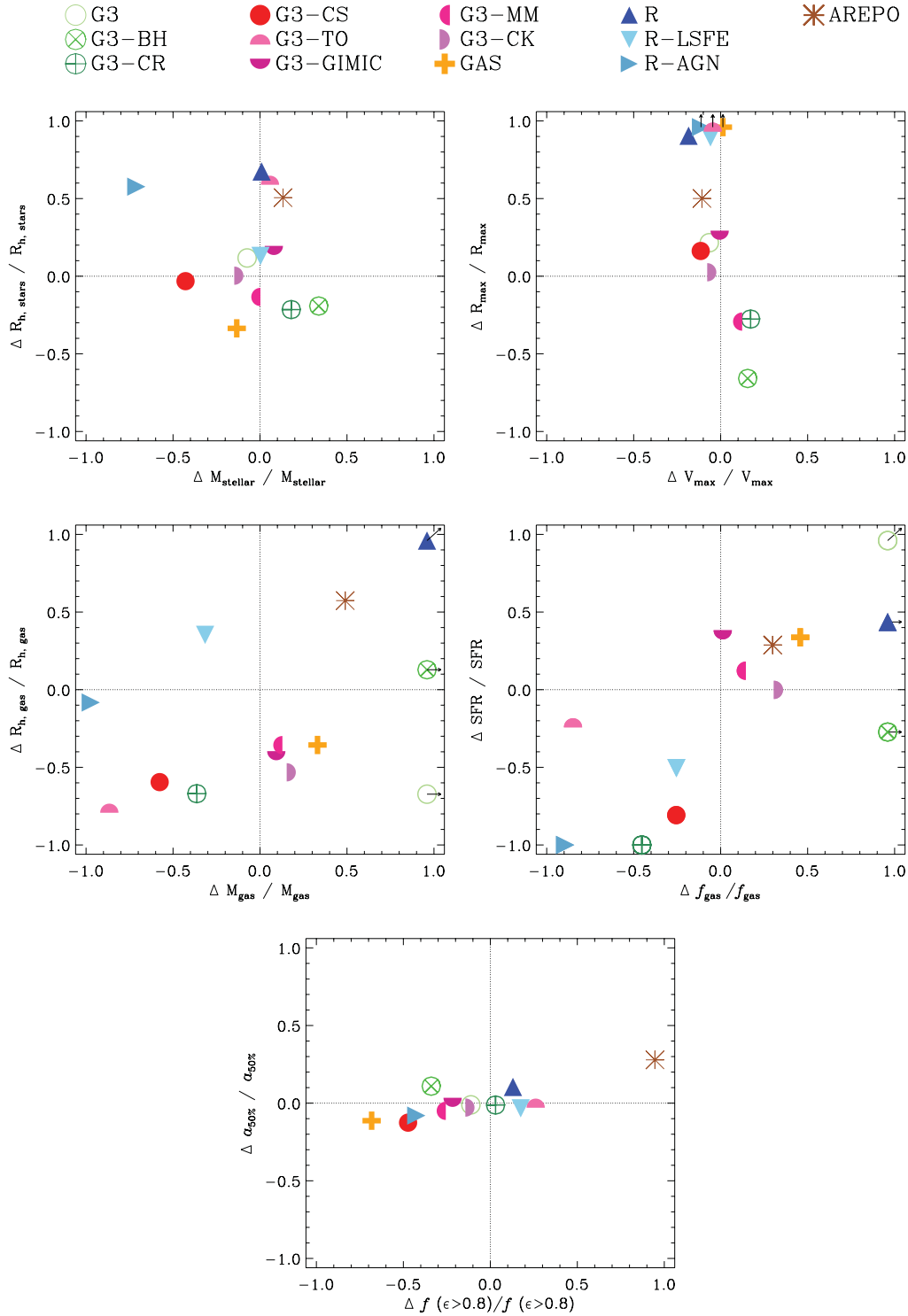
## 4 SUMMARY AND CONCLUSIONS

The Aquila project compares the results of simulations of galaxy formation within a Milky Way-sized  $\Lambda$ CDM halo. We use nine gas-dynamical codes developed and run independently by different groups adopting in each case their preferred implementation of radiative cooling, star formation and feedback. The codes include SPH-based, grid-based and moving mesh techniques; all include SN feedback in thermal and/or kinetic form and primordial or metal-dependent cooling functions. Two of the codes (GADGET and RAMSES) were run three times with three different subgrid physics for a total of 13 different simulations. In addition, each code was run at two different resolution levels in order to investigate numerical convergence. All runs share the same initial conditions and are analysed with a set of consistent analysis tools. Our main conclusions may be summarized as follows.

*Galaxy morphology.* At  $z = 0$ , simulated galaxies exhibit complex morphologies, with spheroids, discs, and bars of varying importance. Morphology shows no obvious dependence on the hydrodynamical method adopted or on numerical resolution, and seems to be mostly related to how and when gas is accreted and transformed into stars. The best indicator of the presence of a disc seems to be the median age of the stars; the later stars form the more prevalent the disc component is. This suggests that, to be successful at forming discs, codes must be able to pre-empt the early transformation of gas into stars while at the same time promoting the accretion and retention of late-accreting, high angular momentum gas.

*Stellar mass.* Despite the common halo-assembly history, we find large code-to-code variations in the final mass of simulated galaxies. The stellar mass ranges from  $4 \times 10^{10}$  to  $\sim 3 \times 10^{11} M_{\odot}$ , depending largely on the adopted strength of the feedback. There is also an indication that the numerical method may play a role: AREPO is able to form almost twice as many stars as G3 although they both adopt the same subgrid physics. Most simulated galaxies are too massive compared with theoretical expectations based on abundance-matching considerations. The median stellar age also correlates with galaxy mass, indicating that models that favour late star formation (as needed to form a disc) do so at the expense of allowing too many stars to form overall.

*Rotation curves.* All simulated galaxies have rotation speeds in excess of what is expected from the Tully–Fisher relation of late-type spirals. The disagreement worsens as the stellar mass of the simulated galaxy increases, both for galaxies with and without a



**Figure 13.** Numerical convergence. We show the fractional variation in various galaxy properties between the level-5 and level-6 resolution runs of each code:  $\Delta Q/Q = (Q_6 - Q_5)/Q_5$ . Results are shown for the stellar mass ( $M_{\text{stellar}}$ ) and half-mass radius ( $R_{\text{h,stars}}$ ); for the peak circular velocity ( $V_{\text{max}}$ ) and corresponding radius ( $r_{\text{max}}$ ); for the gas mass ( $M_{\text{gas}}$ ) and half-mass radius ( $R_{\text{h,gas}}$ ); for the gas fraction ( $f_{\text{gas}}$ ) and present-day SFR and for the median star formation time ( $a_{50\text{ per cent}}$ ) and the fraction of stars with circularities larger than 0.8,  $f(\epsilon > 0.8)$ ). Arrows indicate that results lie outside the plotted range. The actual values of the quantities used for the plot are listed in Table 3.

well-defined stellar disc. At the high-mass end, simulated galaxies have unrealistically sharply peaked, strongly declining rotation curves. Although reasonably flat rotation curves are obtained at low  $M_{\text{stellar}}$ , the Tully–Fisher zero-point offset persists for these systems.

*Galaxy size.* The difficulties matching the Tully–Fisher relation are due to the fact that most simulated galaxies have stellar half-mass radii much smaller than expected from observation given their stellar mass. This is especially true for galaxies where inefficient feedback allows the formation of an overly massive galaxy; these



systems are, quite unrealistically, smaller even than dense early-type galaxies. Galaxies where feedback is able to limit the stellar mass to more acceptable levels are also more concentrated than typical spirals, highlighting the difficulty that all codes face to prevent too many low angular momentum baryons from assembling into the galaxy.

*Star formation history.* The excess of low angular momentum baryons alluded to above may be traced to the inability of feedback schemes to prevent large bursts of star formation at early times. This is clearly seen in the stellar age distribution, which shows that star formation peaks at  $z \sim 4$  and declines steadily afterwards. The same trend holds for essentially all runs, albeit modulated by the particular implementation of feedback adopted by each individual code. Indeed, essentially all models allow more than half of the stars to form in the first  $\sim 3$  Gyr of evolution. The relative insignificance of star formation in recent times compared to the intensity of the early burst seems to be at the root of many of the difficulties that simulated galaxies have in matching observation.

*Gas component.* The properties of the gaseous component of the galaxy at  $z = 0$  show even wider code-to-code variations than the stars, since it is continuously resupplied by accretion, depleted by star formation, and dislodged by feedback-driven winds. This results in large short-lived fluctuations that lead to poor numerical convergence and large code-to-code scatter. Most simulated galaxies have gaseous discs with sizes that compare favourably with observation, although their gas fractions are systematically lower than those of star-forming spirals of comparable mass.

*Numerical convergence.* We have investigated numerical convergence by comparing the results at two different numerical resolutions, spanning a range of  $\sim 2$  and  $\sim 8$  in spatial and mass resolution, respectively. Although numerical convergence is not particularly good for any of the codes, reasonably good convergence is found for the properties of the stellar component, such as total mass and median age. Less well converged are the internal properties of the galaxy, such as the half-mass radius, or the fraction of stars in a rotationally supported disc. For the same reasons cited in the previous item, the properties of the gas are the ones that are most poorly reproduced at the two different resolutions. There is also indication that feedback schemes that are more effective at limiting the stellar mass of the galaxy (such as through the inclusion of AGN-related feedback in addition to SNe) converge less well than other implementations. Overall, the variations introduced by resolution are small compared to code-to-code variations, which leads us to conclude that none of the trends highlighted above is driven primarily by resolution.

*SPH versus AMR versus moving mesh.* Our results suggest that numerical hydrodynamics techniques have some influence on the outcome of a simulation. This is most clearly demonstrated by comparing the results of G3 and AREPO which, despite adopting the same subgrid physics modules, yield galaxies that differ by almost a factor of 2 in stellar mass. The AMR technique also yields large stellar masses (when similarly inefficient feedback is adopted, as in run R), lending support to the view that standard SPH-based codes may underestimate the total amount of gas that can cool and become available for star formation at least in the weak feedback regime. On the other hand, the galaxies formed by R, G3 and AREPO are unrealistically massive and concentrated, so large modifications to the feedback implementation of these codes are needed in order to bring them into agreement with observation. These changes may overwhelm the method-induced differences; for example, R makes a prominent disc while R-AGN has five times fewer stars and no disc. It is therefore unclear at this point whether the shortcomings

of SPH are actually significant compared with the uncertainties involved in designing a star formation/feedback scheme that can yield realistic galaxies, although it is obviously desirable to avoid numerical inaccuracies as far as possible. Further investigation of this question is needed to clarify this issue.

Aside from these considerations, perhaps the main result of the Aquila project is that, despite the large spread in properties spanned by the simulated galaxies, none of them has properties fully consistent with theoretical expectations or observational constraints in terms of mass, size, gas content and morphology. Despite this apparent failure, we believe that the Aquila project nevertheless yields interesting clues to guide how codes might be modified to yield realistic galaxies. For example, the need (i) to control effectively the overcooling of baryons; (ii) to curtail the early burst in star-forming activity and (iii) to promote the accretion and retention of the late-accreting, high angular momentum baryons needed to form discs similar to those of normal spirals are of general applicability to all codes.

There seems to be little predictive power at this point in state-of-the-art simulations of galaxy formation; these seem best suited to the identification of the role and importance of various mechanisms rather than to the detailed modelling of individual systems. It may be argued that the strength of this conclusion depends on whether the parent halo of the Aquila runs (Aq-C) is truly destined to harbour a disc galaxy and that there is no hard proof for this. Further, the possibility that Aq-C might be an unrepresentative outlier should also be considered, as suggested by the L-GALAXIES semi-analytic model (see e.g. Fig. 6).

These objections may be addressed when simulations are able to follow a statistically significant number of galaxies in a cosmologically representative volume. We might not know what kind of galaxy inhabits an individual halo, but we do know what the *population* of galaxies looks like. Evolving a region large enough to contain at least a few dozen Milky-Way-sized galaxies at the resolution achieved here seems like a natural next step, and one that should be achievable in the not too-distant future.

Finally, the complexity of the problem suggests that the best approach to improving galaxy formation simulations may be one where multiple numerical alternatives are developed and explored simultaneously and independently, provided that they are periodically contrasted in controlled experiments such as the one presented here. Given the intricacy of the task and in the absence of a clear front-runner, the diversity of numerical techniques presently available might actually turn out to be a strength rather than a shortcoming.

## ACKNOWLEDGMENTS

We thank the referee for a thorough reading of this work and for his/her prompt answer, helpful comments and suggestions. We are grateful to Aaron Dutton and Marla Geha for providing the compilations of observational data used in this work. We thank Andrew Cooper for providing the predictions of the GALFORM semi-analytic model and for stimulating discussions and suggestions. The G3, G3-BH, G3-CR, G3-CS and AREPO simulations were carried out at the Computing Center of the Max-Planck-Society in Garching. MW acknowledges support by the DFG cluster of excellence ‘Origin and Structure of the Universe’ and VS by the DFG Research Center SFB 881 ‘The Milky Way System’. The RAMSES simulations were performed thanks to the HPC resources of CINES under the allocation 2010-GEN2192 made by GENCI, France. PM and GM acknowledge a *standard 2010 grant* from CASPUR, and support

from PRIN-INAF 2009 titled ‘Towards an Italian Network for Computational Cosmology’. This work was supported by Marie Curie Initial Training Network CosmoComp (PITN-GA-2009-238356). The GASOLINE simulations used computers graciously provided by SHARCNET, Compute Canada and the HPCAVF at the University of Central Lancashire in Preston, UK. The G3-TO simulations were performed with T2K-Tsukuba at the Center for Computational Sciences at the University of Tsukuba. TO acknowledges financial support from Grant-in-Aid for Young Scientists (start-up: 21840015) and from MEXT HPCI Strategic Program. Some of the calculations for this paper were performed on the Green Machine at Swinburne University of Technology and on the ICC Cosmology Machine, which is part of the DiRAC Facility, jointly funded by STFC, the Large Facilities Capital Fund of BIS and Durham University. We thank the DEISA Consortium ([www.deisa.eu](http://www.deisa.eu)), co-funded through the EU FP6 project RI-031513 and the FP7 project RI-222919, for support within the DEISA Extreme Computing Initiative. TQ and EC acknowledge support from NSF grant AST-0908499.

## REFERENCES

- Abadi M. G., Navarro J. F., Steinmetz M., Eke V. R., 2003a, *ApJ*, 591, 499  
 Abadi M. G., Navarro J. F., Steinmetz M., Eke V. R., 2003b, *ApJ*, 597, 21  
 Agertz O. et al., 2007, *MNRAS*, 380, 963  
 Agertz O., Teyssier R., Moore B., 2011, *MNRAS*, 410, 1391  
 Bell E. F., de Jong R. S., 2000, *MNRAS*, 312, 497  
 Blitz L., Rosolowsky E., 2006, *ApJ*, 650, 933  
 Booth C. M., Schaye J., 2009, *MNRAS*, 398, 53  
 Booth C. M., Schaye J., 2010, *MNRAS*, 405, L1  
 Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645  
 Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, *MNRAS*, 398, 1150  
 Brook C. B. et al., 2011, *MNRAS*, 415, 1051  
 Brooks A. M. et al., 2011, *ApJ*, 728, 51  
 Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000  
 Bryan G. L., Norman M. L., 1995, *BAAS*, 27, 1421  
 Cen R., Ostriker J., 1992, *ApJ*, 393, 22  
 Ceverino D., Klypin A., 2009, *ApJ*, 695, 292  
 Chabrier G., 2003, *PASP*, 115, 763  
 Colin P., Avila-Reese V., Vázquez-Semadeni E., Valenzuela O., Ceverino D., 2010, *ApJ*, 713, 535  
 Cooper A. P. et al., 2010, *MNRAS*, 406, 744  
 Courteau S., Dutton A. A., van den Bosch F. C., MacArthur L. A., Dekel A., McIntosh D. H., Dale D. A., 2007, *ApJ*, 671, 203  
 Crain R. A. et al., 2009, *MNRAS*, 399, 1773  
 Crain R. A., McCarthy I. G., Frenk C. S., Theuns T., Schaye J., 2010, *MNRAS*, 407, 1403  
 Creasey P., Theuns T., Bower R. G., Lacey C. G., 2011, *MNRAS*, 415, 3706  
 Cui W., Springel V., Yang X., De Lucia G., Borgani S., 2011, *MNRAS*, 416, 2997  
 Dalcanton J. J., Spergel D. N., Summers F. J., 1997, *ApJ*, 482, 659  
 Dalla Vecchia C., Schaye J., 2008, *MNRAS*, 387, 1431  
 Deason A. J. et al., 2011, *MNRAS*, 415, 2607  
 Dehnen W., McLaughlin D. E., Sachania J., 2006, *MNRAS*, 369, 1688  
 De Lucia G., Blaizot J., 2007, *MNRAS*, 375, 2  
 Di Matteo T., Croft R. A. C., Springel V., Hernquist L., 2003, *ApJ*, 593, 56  
 Dubois Y., Teyssier R., 2008, *A&A*, 477, 79  
 Dutton A. A. et al., 2011, *MNRAS*, 416, 322  
 Fabjan D., Borgani S., Tornatore L., Saro A., Murante G., Dolag K., 2010, *MNRAS*, 401, 1670  
 Fall S. M., Efstathiou G., 1980, *MNRAS*, 193, 189  
 Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, *PASP*, 110, 761  
 Font A. S., McCarthy I. G., Crain R. A., Theuns T., Schaye J., Wiersma R. P. C., Dalla Vecchia C., 2011, *MNRAS*, 416, 2802  
 Fryxell B. et al., 2000, *ApJS*, 131, 273  
 Gingold R. A., Monaghan J. J., 1977, *MNRAS*, 181, 375  
 Governato F. et al., 2004, *ApJ*, 607, 688  
 Governato F., Willman B., Mayer L., Brooks A., Stinson G., Valenzuela O., Wadsley J., Quinn T., 2007, *MNRAS*, 374, 1479  
 Governato F. et al., 2009, *MNRAS*, 398, 312  
 Governato F. et al., 2010, *Nat*, 463, 203  
 Guedes J., Callegari S., Madau P., Mayer L., 2011, *ApJ*, 742, 76  
 Guo Q., White S., Li C., Boylan-Kolchin M., 2010, *MNRAS*, 404, 1111  
 Guo Q. et al., 2011, *MNRAS*, 413, 101  
 Haardt F., Madau P., 1996, *ApJ*, 461, 20  
 Haardt F., Madau P., 2001, in Neumann D. M., Tran J. T. V., eds, *Clusters of Galaxies and the High Redshift Universe Observed in X-rays*. Moriond Astrophysics Meeting, France  
 Haynes M. P., Giovanelli R., Chamaux P., da Costa L. N., Freudling W., Salzer J. J., Wegner G., 1999, *AJ*, 117, 2039  
 Joung M. R., Cen R., Bryan G. L., 2009, *ApJ*, 692, L1  
 Jubelgas M., Springel V., Ensslin T., Pfrommer C., 2008, *A&A*, 481, 33  
 Katz N., 1992, *ApJ*, 391, 502  
 Katz N., Weinberg D. H., Hernquist L., 1996, *ApJS*, 105, 19  
 Kauffmann G., White S. D. M., Guiderdoni B., 1993, *MNRAS*, 264, 201  
 Kennicutt R. C., Jr, 1998, *ApJ*, 498, 541  
 Keres D., Vogelsberger M., Sijacki D., Springel V., Hernquist L., 2011, *arXiv e-prints (arXiv:1109.4638)*  
 Kobayashi C., 2004, *MNRAS*, 347, 740  
 Kobayashi C., Nakasato N., 2011, *ApJ*, 729, 16  
 Kobayashi C., Nomoto K., 2009, *ApJ*, 707, 1466  
 Kobayashi C., Tsujimoto T., Nomoto K., Hachisu I., Kato M., 1998, *ApJ*, 503, L155  
 Kobayashi C., Umeda H., Nomoto K., Tominaga N., Ohkubo T., 2006, *ApJ*, 653, 1145  
 Kobayashi C., Springel V., White S. D. M., 2007, *MNRAS*, 376, 1465  
 Kobayashi C., Karakas A. I., Umeda H., 2011, *MNRAS*, 414, 3231  
 Kravtsov A. V., 1999, PhD thesis, New Mexico State Univ.  
 Kroupa P., Tout C. A., Gilmore G., 1993, *MNRAS*, 262, 545  
 Leitner S. N., Kravtsov A. V., 2011, *ApJ*, 734, 48  
 Li Y., White S. D. M., 2008, *MNRAS*, 384, 1459  
 Lucy L. B., 1977, *AJ*, 82, 1013  
 McCarthy I. G., Font A. S., Crain R. A., Deason A. J., Schaye J., Theuns T., 2011, *MNRAS*, 419, 3304  
 McKee C. F., Ostriker J. P., 1977, *ApJ*, 218, 148  
 Marigo P., 2001, *A&A*, 370, 194  
 Mayer L., Governato F., Kaufmann T., 2008, *Advanced Sci. Lett.*, 1, 7  
 Mitchell N. L., McCarthy I. G., Bower R. G., Theuns T., Crain R. A., 2009, *MNRAS*, 395, 180  
 Mo H. J., Mao S., White S. D. M., 1998, *MNRAS*, 295, 319  
 Monaco P., Murante G., Borgani S., Dolag K., 2012, *MNRAS*, 421, 2485  
 Monaghan J. J., 1992, *ARA&A*, 30, 543  
 Murante G., Monaco P., Giovali M., Borgani S., Diaferio A., 2010, *MNRAS*, 405, 1491  
 Navarro J. F., Benz W., 1991, *ApJ*, 380, 320  
 Navarro J. F., Steinmetz M., 1997, *ApJ*, 478, 13  
 Navarro J. F., Steinmetz M., 2000, *ApJ*, 538, 477  
 Navarro J. F., Frenk C. S., White S. D. M., 1995, *MNRAS*, 275, 56  
 Okabe A., Boots B., Sugihara K., Nok-Chiu S., ed., 2000, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. John Wiley & Sons Ltd, Chichester  
 Okamoto T., Frenk C. S., 2009, *MNRAS*, 399, L174  
 Okamoto T., Jenkins A., Eke V. R., Quilis V., Frenk C. S., 2003, *MNRAS*, 345, 429  
 Okamoto T., Eke V. R., Frenk C. S., Jenkins A., 2005, *MNRAS*, 363, 1299  
 Okamoto T., Nemmen R. S., Bower R. G., 2008, *MNRAS*, 385, 161  
 Okamoto T., Frenk C. S., Jenkins A., Theuns T., 2010, *MNRAS*, 406, 208  
 Oliver S. et al., 2010, *MNRAS*, 405, 2279  
 Parry O. H., Eke V. R., Frenk C. S., Okamoto T., 2012, *MNRAS*, 419, 3304  
 Piontek F., Steinmetz M., 2011, *MNRAS*, 410, 2625  
 Pizagno J. et al., 2007, *AJ*, 134, 945  
 Portinari L., Chiosi C., Bressan A., 1998, *A&A*, 334, 505

- Quilis V., 2004, MNRAS, 352, 1426
- Rasera Y., Teyssier R., 2006, A&A, 445, 1
- Sales L. V., Navarro J. F., Schaye J., Dalla Vecchia C., Springel V., Booth C. M., 2010, MNRAS, 409, 1541
- Sawala T., Guo Q., Scannapieco C., Jenkins A., White S., 2011, MNRAS, 413, 659
- Scannapieco C., Tissera P. B., White S. D. M., Springel V., 2005, MNRAS, 364, 552
- Scannapieco C., Tissera P. B., White S. D. M., Springel V., 2006, MNRAS, 371, 1125
- Scannapieco C., Tissera P. B., White S. D. M., Springel V., 2008, MNRAS, 389, 1137
- Scannapieco C., White S. D. M., Springel V., Tissera P. B., 2009, MNRAS, 396, 696
- Scannapieco C., Gadotti D. A., Jonsson P., White S. D. M., 2010, MNRAS, 407, L41
- Scannapieco C., White S. D. M., Springel V., Tissera P. B., 2011, MNRAS, 417, 154
- Schaye J., 2004, ApJ, 609, 667
- Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210
- Schaye J. et al., 2010, MNRAS, 402, 1536
- Schombert J. M., McGaugh S. S., Eder J. A., 2001, AJ, 121, 2420
- Shen S., Wadsley J., Stinson G., 2010, MNRAS, 407, 1581
- Sijacki D., Vogelsberger M., Keres D., Springel V., Hernquist L., 2011, arXiv e-prints (arXiv:1109.3468)
- Smith M. C. et al., 2007, MNRAS, 379, 755
- Sofue Y., Honma M., Omodaka T., 2009, PASJ, 61, 227
- Sommer-Larsen J., Dolgov A., 2001, ApJ, 551, 608
- Sommer-Larsen J., Gelato S., Vedel H., 1999, ApJ, 519, 501
- Springel V., 2005, MNRAS, 364, 1105
- Springel V., 2010a, MNRAS, 401, 791
- Springel V., 2010b, ARA&A, 48, 391
- Springel V., Hernquist L., 2003, MNRAS, 339, 289
- Springel V., Di Matteo T., Hernquist L., 2005a, MNRAS, 361, 776
- Springel V. et al., 2005b, Nat, 435, 629
- Springel V., Frenk C. S., White S. D. M., 2006, Nat, 440, 1137
- Springel V. et al., 2008, MNRAS, 391, 1685
- Stinson G., Seth A., Katz N., Wadsley J., Governato F., Quinn T., 2006, MNRAS, 373, 1074
- Stinson G. S., Bailin J., Couchman H., Wadsley J., Shen S., Nickerson S., Brook C., Quinn T., 2010, MNRAS, 408, 812
- Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253
- Swaters R. A., Schoenmakers R. H. M., Sancisi R., van Albada T. S., 1999, MNRAS, 304, 330
- Tasker E. J., Brunino R., Mitchell N. L., Michielsen D., Hopton S., Pearce F. R., Bryan G. L., Theuns T., 2008, MNRAS, 390, 1267
- Teyssier R., 2002, A&A, 385, 337
- Teyssier R., Moore B., Martizzi D., Dubois Y., Mayer L., 2011, MNRAS, 414, 195
- Torrey P., Vogelsberger M., Sijacki D., Springel V., Hernquist L., 2011, arXiv e-prints (arXiv:1110.5635)
- van den Bosch F. C., Abel T., Croft R. A. C., Hernquist L., White S. D. M., 2002, ApJ, 576, 21
- Verheijen M. A. W., 2001, ApJ, 563, 694
- Vogelsberger M., Sijacki D., Keres D., Springel V., Hernquist L., 2011, arXiv e-prints (arXiv:1109.1281)
- Wadepuhl M., Springel V., 2011, MNRAS, 410, 1975
- Wadsley J. W., Stadel J., Quinn T., 2004, New Astron., 9, 137
- Wadsley J. W., Veeravalli G., Couchman H. M. P., 2008, MNRAS, 387, 427
- Watkins L. L., Evans N. W., An J. H., 2010, MNRAS, 406, 264
- Weil M. L., Eke V. R., Efstathiou G., 1998, MNRAS, 300, 773
- Wiersma R. P. C., Schaye J., Smith B. D., 2009a, MNRAS, 393, 99
- Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, MNRAS, 399, 574
- Xue X. X. et al., 2008, ApJ, 684, 1143
- Zavala J., Okamoto T., Frenk C. S., 2008, MNRAS, 387, 364

## APPENDIX A: DESCRIPTION OF THE CODES

In this appendix we summarize the main properties of the various simulation codes and implemented physics. These succinct descriptions have been provided by each of the individual groups participating in the Aquila project. As explained below, the different models use a variety of implementations of the processes of star formation and feedback, resulting in the SFR surface densities shown in Fig. A1.

### A1 GADGET-3 models (G3, G3-BH, G3-CR)

The simulations G3, G3-BH and G3-CR are based on GADGET3. This is a significantly updated version of GADGET2 (Springel 2005), a fully cosmological code based on SPH and on the TREE-PM method to evaluate gravitational forces. All three of these simulations use the star formation model introduced by Springel & Hernquist (2003). G3-BH includes AGN feedback following Springel et al. (2005a), while the G3-CR model includes, in addition, energy deposition by CRs (see Jubelgas et al. 2008, and references therein).

The Springel & Hernquist (2003) star formation model uses a primordial cooling function following Katz et al. (1996) and includes a uniform UV background, based on an updated version of Haardt & Madau (1996), that is switched on at  $z = 6$ . Stars are formed stochastically in dense regions assuming a Salpeter IMF. This model pictures each gas element as a two-phase mixture of hot and cold gas in approximate pressure equilibrium. Cold gas is converted into ‘star’ particles on a characteristic time-scale  $t_*$ . Assuming that a fraction  $\beta$  of the stellar population represented by each star particle is so short-lived that they explode as SNe instantly, their (metal-enriched) mass is fed back to the interstellar medium. The SFR is given by

$$\frac{d\rho_*}{dt} = \frac{\rho_c}{t_*} - \beta \frac{\rho_c}{t_*} = (1 - \beta) \frac{\rho_c}{t_*}, \quad (\text{A1})$$

where  $\rho_c$  is the mean mass density of gas in the cold phase, and  $t_*$  is a density-dependent star formation time-scale. The energy feedback of SNe explosions is used to heat the ambient hot gas as well as to evaporate cold clouds, leading to a self-regulating cycle for star-forming gas. The net effect is that the dynamics of the multiphase medium is governed by an effective equation-of-state (see Springel & Hernquist 2003).

The simulations G3-BH and G3-CR both include thermal AGN feedback associated with a Bondi–Hoyle–Lyttelton parametrization of the (spatially unresolved) gas accretion on to a central SMBH (see Springel et al. 2005a, for a detailed description). The thermal feedback energy is parametrized by  $\dot{E}_{\text{feed}} = \epsilon_f \epsilon_r \dot{M}_{\text{BH}} c^2$ , where  $\epsilon_r$  is the radiative efficiency of the black hole and  $\epsilon_f$  encodes the feedback efficiency.

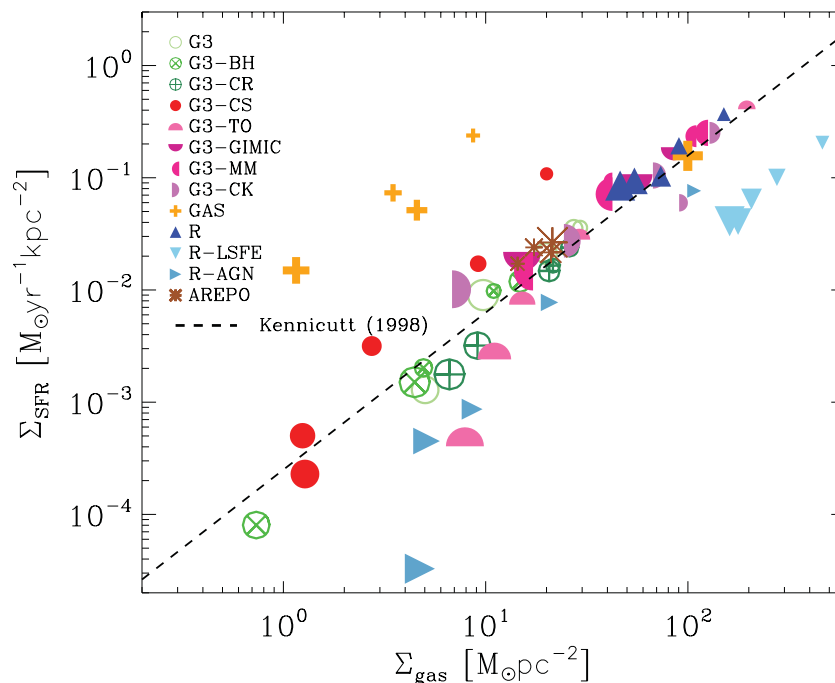
In G3-CR the distribution function of CRs is modelled as a power law,

$$\frac{d^2 N}{dp dv} = C p^{-\alpha} \theta(p - q), \quad (\text{A2})$$

in momentum space. The non-thermal pressure of this CR population is then given by

$$P_{\text{CR}} = \frac{C m_p C^2}{6} \mathcal{B}_{\frac{1}{1+q^2}} \left( \frac{\alpha - 2}{2}, \frac{3 - \alpha}{2} \right), \quad (\text{A3})$$

which is added to the ordinary gas pressure (Jubelgas et al. 2008). Here  $\mathcal{B}_n(a, b)$  denotes incomplete  $\beta$  functions and  $\alpha$  is the assumed power-law slope of the CR particles. The energy injected by SN



**Figure A1.** Cold ( $T < 10^5$  K) gas surface density versus star formation rate per unit area in the simulated galaxies, measured face-on and averaged within a cylinder of radius equal to the (projected) half-mass radius of the cold gas (see the right-hand panel of Fig. 9). Each set of symbols indicate results for level-5 simulations at  $z = 2, 1.5, 1, 0.5$  and  $0$ , with the size of the symbols increasing with decreasing redshift. The dashed line indicates, for reference, the Schmidt–Kennicutt law for nearby ‘normal’ and ‘star-bursting’ discs, as given by Kennicutt (1998).

explosions into the CR population is given as  $\Delta \tilde{\epsilon}_{\text{SN}} = \zeta_{\text{SN}} \epsilon_{\text{SN}} \dot{m}_* \Delta t$ , where  $\zeta_{\text{SN}}$  is the fraction of the SN energy that first appears in CRs.

The numerical parameters used for the Aquila project simulations are as follows: a star formation time-scale of  $t_* = 2.1$  Gyr and a SN energy of  $E_{\text{SN}} = 10^{51}$  erg. The G3-BH model adopts  $\epsilon_r = 0.1$  and  $\epsilon_f = 0.05$ . The G3-CR run uses a CR generation efficiency of  $\zeta_{\text{SN}} = 0.3$  and a spectral index  $\alpha = 2.5$ .

## A2 The CS model (G3-CS)

The CS model is a GADGET3-based code that includes stochastic star formation, chemical enrichment and SN (thermal) feedback from Type II and Type Ia SN explosions, a multiphase model for the gas component and metal-dependent cooling. Full details of the implementation are given in Scannapieco et al. (2005) and Scannapieco et al. (2006) and previous applications to cosmological galaxy formation and disc formation can be found in Scannapieco et al. (2008, 2009, 2010, 2011).

The model includes a UV background field which turns on at  $z = 6$  and follows the formulation of Haardt & Madau (1996) and metal-dependent cooling functions from Sutherland & Dopita (1993) are used. Star formation occurs stochastically in gas particles with high density ( $n > n_{\text{th}}$ ) and in a convergent flow (Springel & Hernquist 2003), following the Schmidt–Kennicutt law:

$$\frac{d\rho_*}{dt} = c_* \frac{\rho_{\text{gas}}}{t_{\text{dyn}}}, \quad (\text{A4})$$

where  $\rho_*$  and  $\rho_{\text{gas}}$  are the stellar and gas densities,  $c_*$  is a star formation efficiency and  $t_{\text{dyn}}$  is the dynamical time of the gas. Stars return metals and energy during SN Type II and Type Ia explosions. Both types of SNe eject the same amount of energy  $E_{\text{SN}}$  to the interstellar medium, but they have different chemical yields, explosion times and rates. SN energy is dumped in given fractions to the cold ( $\epsilon_c$ )

and hot ( $1 - \epsilon_c$ ) neighbours of exploding stars. Energy deposition into hot neighbours occurs at the time of explosion; however, for cold neighbours energy from successive explosions is instead accumulated and deposited only after a time-delay which depends on the local conditions of the cold and hot gas phases. In this way, artificial loss of SN energy in high-density regions is prevented. Star particles are treated as single stellar populations with a Salpeter IMF.

We introduce a multiphase gas model which allows gas in both dense and diffuse phases to co-exist in the same spatial region. In our model, SPH particles of a given physical state (density, entropy) ignore neighbouring particles that have a much lower (a factor of 50) entropy. This scheme also makes the deposition of SN energy more efficient.

The input parameters used for the simulations analysed in this paper are  $n_{\text{th}} = 0.03 \text{ cm}^{-3}$ ,  $c_* = 0.1$ ,  $E_{\text{SN}} = 0.7 \times 10^{51}$  erg and  $\epsilon_c = 0.5$ . Finally, the input parameters for the chemical model are identical to those used in Scannapieco et al. (2009).

## A3 The TO model (G3-TO)

The TO model is described in Okamoto et al. (2010) and is based on an early version of the GADGET3 code. It incorporates metal-dependent cooling, star formation, thermal and kinetic feedback from SNe and enrichment by asymptotic giant branch (AGB) stars and SNe. Examples of its application in cosmological galaxy formation simulations include Okamoto & Frenk (2009), Okamoto et al. (2010) and Parry et al. (2012).

Photoheating and cooling are implemented as described in Wiersma et al. (2009a), including contributions from 11 elements in the presence of a spatially uniform, time evolving UV background of the form calculated by Haardt & Madau (2001). Gas above a density  $n_{\text{th}}$  forms stars at a rate normalized to reproduce the Schmidt–Kennicutt law. Star particles represent simple stellar



populations (SSPs) with a Chabrier (2003) IMF. Following Wiersma et al. (2009b), energy, mass and metals are returned to the ISM by AGB stars, Type Ia and Type II SNe on time-scales appropriate for the age and metallicity of the stellar population, with yields and stellar lifetimes taken from Portinari, Chiosi & Bressan (1998) and Marigo (2001).

The unresolved ISM is modelled using a subgrid prescription. Each gas particle with  $n > n_{\text{th}}$  is treated as a series of cold clouds, with an empirically motivated mass function, embedded in a hot ambient medium. The two phases exchange mass through thermal instability and cloud evaporation. Each cloud has a SFR that is inversely proportional to its dynamical time. The hot phase is supported by imposing a minimum pressure of  $P_{\text{min}} \propto \rho^{1.4}$ . Type Ia SNe increase the thermal energy of the gas around star-forming regions, whilst Type II SNe are assumed to drive large-scale winds. The wind speed is chosen to be proportional to the halo circular velocity, using the local dark matter velocity dispersion as a proxy ( $v = \alpha \sigma_{\text{DM}}$ ). The mass loading then follows by requiring conservation of all of the available SNe energy. Wind particles are decoupled from hydrodynamic forces for a short time, to allow them to escape the star-forming region and ensure that the specified mass loading and wind velocity are not modified by viscous drag from the ISM.

The density required for star formation is set at  $n_{\text{th}} = 0.1 \text{ cm}^{-3}$  in the level-6 simulation and  $0.4 \text{ cm}^{-3}$  in the level-5 simulation. For the wind speed parameter,  $\alpha = 5$  is used, which has been shown to produce a good match to the Milky Way satellite luminosity function (Okamoto et al. 2010; Parry et al. 2012).

#### A4 The GIMIC model (G3-GIMIC)

The GIMIC model is a GADGET3-based code that includes metal-dependent cooling on an element-by-element basis in the presence of a UV background, star formation and SNe-driven winds, as well as mass and metal recycling by AGB stars, Type Ia and Type II SNe. GIMIC is identical to model MILL of the OWLS suite of simulations (Schaye et al. 2010). A full description of the model can be found in Crain et al. (2009) and Schaye et al. (2010) and further applications of this model can be found in Crain et al. (2010), Cui et al. (2011), Deason et al. (2011), Font et al. (2011) and McCarthy et al. (2011).

The model includes a spatially uniform, time evolving UV background following the formulation of Haardt & Madau (2001). Hydrogen reionizes at  $z = 9$ , He II at  $z = 3.5$ . Radiative cooling and heating processes are implemented on an element-by-element basis, using interpolation tables computed with CLOUDY (Ferland et al. 1998), as described by Wiersma et al. (2009a). The SFR prescription, described by Schaye & Dalla Vecchia (2008), is pressure dependent and enforces a local Schmidt–Kennicutt law and requires only that the slope and normalization of the observed relation are specified as input parameters. Gas particles are converted to star particles stochastically, with a probability that depends on their associated SFR. Each star particle represents a SSP with a Chabrier (2003) IMF and inherits the elemental abundances of its parent gas particle. The chemodynamical evolution of SSPs and the associated recycling of heavy elements into surrounding gas is followed on an element-by-element basis and includes contributions from AGB stars and both Type Ia and Type II SNe, as described by Wiersma et al. (2009b).

GIMIC appeals to a phenomenological treatment to model the energetic feedback resulting from stellar evolution and SNe. As the simulation lacks both the physics and the resolution to model the multiphase ISM, an effective equation of state is imposed on to

gas particles that are sufficiently dense ( $n_{\text{H}} > 0.1 \text{ cm}^{-3}$ ) and cold ( $T < 10^5 \text{ K}$ ) to be subject to gravitational instability (Schaye 2004). The effective equation of state,  $P = \kappa \rho^{4/3}$ , is chosen to ensure that the Jeans mass and the ratio between the Jeans length and the SPH smoothing length are independent of the gas density, thus preventing spurious fragmentation due to a lack of numerical resolution (Schaye & Dalla Vecchia 2008). SN energy is assumed to drive galaxy-wide outflows, therefore, gas particles neighbouring star-forming regions are given a randomly orientated velocity kick of  $600 \text{ km s}^{-1}$ . A probabilistic scheme ensures that, on average, the mass put into the wind is a factor of 4 times the amount of stars formed. Unlike many similar schemes, these particles are not temporarily decoupled from hydrodynamic forces (for further discussion, see Dalla Vecchia & Schaye 2008).

Values for all model parameters can be found in Crain et al. (2009), but to summarize: the initial wind mass loading and velocity are set to 4 and  $600 \text{ km s}^{-1}$ , respectively (with  $10^{51} \text{ erg SN}^{-1}$ ), the star formation threshold to  $n > 0.1 \text{ cm}^{-3}$  and the input star formation law to have slope 1.4 and normalization coefficient  $\sim 1.5 \times 10^{-4} \text{ M}_{\odot} \text{ yr}^{-1} \text{ kpc}^{-2}$ .

#### A5 The MM model (G3-MM)

The MM model (or Multi Phase Particle Integrator, MUPPI), fully described in Murante et al. (2010), is also implemented within GADGET3. We include radiative cooling of a gas with primordial composition; heating from a uniform UV background of the form given by Haardt & Madau (1996), turned on at a redshift  $z = 6$ ; star formation and stellar feedback with the multiphase model described below. We assume a Salpeter IMF and the instantaneous recycling approximation to treat gas restoration and SN energy feedback. No chemical enrichment is included in the present version of the code. No treatment of accreting black holes or CR feedback is implemented.

In the MUPPI subresolution model of star formation and stellar feedback, gas particles at relatively high density ( $n > n_{\text{th}}$ ) and low temperature ( $T < T_{\text{th}}$ ) are treated as a multiphase system, made up by cold and hot gas phases co-existing in pressure equilibrium, and a stellar component. Hot phase has low-mass fraction but high filling factor, so its cooling time is much longer than that computed for the average particle density, and thus thermal energy is *not* quickly dissipated as soon as it is injected. Regarding the cold phase, a part of it is assumed to be in molecular form. Blitz & Rosolowsky (2006) found a phenomenological correlation between the ratio of molecular and H I gas surface densities and the external disc pressure. Inspired by this result, we compute the fraction of cold gas in molecular form as a function of hydrodynamical pressure  $P$ :

$$f_{\text{mol}} = \frac{1}{1 + \frac{P_0}{P}}. \quad (\text{A5})$$

Mass flows among the components as follows. Cooling of the hot phase feeds the cold gas phase. Stars form from the *molecular* cold gas:

$$\dot{M}_{\star} = f_{\star} f_{\text{mol}} \frac{M_{\text{cold}}}{t_{\text{dyn}}}, \quad (\text{A6})$$

where  $t_{\text{dyn}}$  is the dynamical time of the cold phase at the onset of a multiphase cycle (see Murante et al. 2010 for details) and  $f_{\star}$  is the fraction of molecular cloud transformed into stars per dynamical time. Our prescription for star formation is not based on imposing a Schmidt–Kennicutt relation, which is instead naturally produced

by our model (see Monaco et al. 2012). A fraction  $f_{\text{ev}}$  of the star-forming gas is evaporated back to the hot phase, while a fraction  $f_{\text{re}}$  of stellar mass is instantaneously restored to the hot phase. The code also tracks the hot phase thermal energy. This gained or lost through hydrodynamics, lost by cooling and gained from Type II SN explosions: a small fraction  $f_{\text{in}}$  of their energy is deposited into the hot phase of the same star-forming gas particle, while a more significant fraction  $f_{\text{out}}$  is distributed to neighbouring particles, preferentially along the ‘least resistance path’ defined as a cone of semi-aperture  $\theta$  and anti-aligned with the density gradient. Finally, to mimic the destruction of molecular clouds a multiphase particle returns single phase after a time  $t_{\text{clock}}$ , scaled with the dynamical time.

A system of ordinary differential equations evolves the mass and energy flows described above. This is solved on-the-fly within each SPH time-step with a Runge–Kutta integrator with adaptive time-step. This dynamical system has an intrinsically runaway behaviour: energy from SNe increases gas pressure, which in turn increases SFR through the higher molecular fraction. This runaway is stabilized by the hydrodynamic response of gas: when a particle receives enough energy, it expands thereby decreasing its pressure. The equations MM model solves are similar to those used in the star formation model by Springel & Hernquist (2003); the main differences are that in MM model no equilibrium solution is assumed, and the effect of hydrodynamics on the multiphase gas is explicitly taken into account.

MUPPI produces reservoirs of ‘virtual’ stars that are transformed into star particles using the stochastic algorithm of Springel & Hernquist (2003). Each gas particle produces up to four generations of star particles.

We used here the same ‘standard’ set of parameters described in Murante et al. (2010), namely: (i) star formation efficiency  $f_* = 0.02$ , amount of molecular gas which is converted into stars; (ii)  $P_0 = 35\,000\text{ K cm}^{-3}$ , pressure normalization for the Blitz & Rosolowsky relation; (iii)  $T_c = 1000\text{ K}$ , temperature of the cold phase; (iv)  $f_{\text{out}} = 0.3$ , fraction of SNe energy given to neighbouring gas particles; (v)  $f_{\text{in}} = 0.02$ , fraction given to the hot gas of the particle itself; (vi)  $f_{\text{ev}} = 0.1$ , fraction of cold gas mass evaporated by SNe; (vii)  $E_{\text{SN}} = 10^{51}\text{ erg}$ , SN energy; (viii)  $\theta = 60^\circ$ , semi-aperture of the cone determining the neighbouring gas particles which receive energy; (ix)  $n_{\text{th}} = 0.01\text{ cm}^{-3}$ , density threshold for entering the multiphase stage; (x)  $T_{\text{th}} = 50\,000\text{ K}$ , temperature threshold for entering the multiphase stage;  $t_{\text{clock}} = 2t_{\text{dyn}}$ , time after which a particle exits the multiphase stage.

## A6 The CK model (G3-CK)

The CK model is a GADGET3-based code that includes star formation, chemical enrichment and (thermal) feedback from stellar winds, core-collapse SNe, Type Ia SNe and AGB stars. The details are described in Kobayashi (2004), Kobayashi et al. (2007) and Kobayashi & Nakasato (2011).

The UV background radiation is included with Haardt & Madau (1996) from  $z = 6$ . Radiative cooling is computed with the metal-dependent cooling functions (Sutherland & Dopita 1993) as a function of  $[\text{Fe}/\text{H}]$ .  $[\text{O}/\text{Fe}]$  is fixed with the observed  $[\text{O}/\text{Fe}]$ – $[\text{Fe}/\text{H}]$  relation in the solar neighbourhood. The star formation criteria are (1) converging flow, (2) rapid cooling and (3) Jeans unstable. The SFR is determined from the Schmidt–Kennicutt law (equation A4). If a gas particle satisfies the star formation criteria, a part of the mass of the gas particle turns into a new star particle. We then treat the star particle as a SSP and calculate the evolution of the stellar

population every time-step. The masses of the stars associated with each star particle are distributed according to an IMF. We adopt the Salpeter IMF that is invariant to time and metallicity with a slope  $x = 1.35$  for  $0.1\text{--}120\text{ M}_\odot$ , to be consistent with the Galactic chemical evolution (Kobayashi et al. 2006).

For the feedback of energy and heavy elements, we do not adopt the instantaneous recycling approximation. Instead, via stellar winds, core-collapse SNe, Type Ia SNe and AGB stars, thermal energy and heavy elements are ejected from an evolved star particle as a function of time, and distributed to a constant number  $N_{\text{FB}}$  of surrounding gas particles. Among core-collapse SNe, we include the effect of hypernovae, which are observationally known to produce more than 10 times larger explosion energy and iron than normal Type II SNe. We adopt the metal-dependent efficiency of hypernovae ( $\epsilon_{\text{HN}} = 0.5, 0.5, 0.4, 0.01$  and  $0.01$  for  $Z = 0, 0.001, 0.004, 0.02$  and  $0.05$ ) for the initial masses of  $M \gtrsim 20\text{ M}_\odot$ , to be consistent with the observed  $[\text{Zn}/\text{Fe}]$  ratios in the Milky Way Galaxy (Kobayashi & Nakasato 2011). The ejected energy ( $1\text{--}40 \times 10^{51}\text{ erg}$ ) and nucleosynthesis yields are taken from Kobayashi et al. (2006) as a function of progenitor mass and metallicity. With hypernovae, cosmological simulations give a better agreement with observed cosmic SFRs (Kobayashi et al. 2007). For Type Ia SNe, we use our single-degenerate model with the metallicity effect (Kobayashi et al. 1998; Kobayashi & Nomoto 2009). The lifetimes span a range of  $0.1\text{--}20\text{ Gyr}$  as a function of progenitor metallicity, which is consistent with the observed SN rates and with the observed  $[\alpha/\text{Fe}]$  relations in the Milky Way Galaxy. The ejected energy is  $1.3 \times 10^{51}\text{ erg}$  per explosion. From stellar winds,  $\sim 0.2 \times 10^{51}\text{ erg}$  is ejected depending on metallicity for the stars with  $\geq 8\text{ M}_\odot$ . The adopted nucleosynthesis yields of SNe and AGB stars are summarized in Kobayashi, Karakas & Umeda (2011).

The input parameters used for the simulations analysed in this paper are the star formation time-scale of  $c_* = 0.02$  and the feedback number of  $N_{\text{FB}} = 64$ .

## A7 The GASOLINE model (GAS)

The GAS model uses the SPH code GASOLINE, which is described in detail in Wadsley et al. (2004). It includes a UV background heating, metal cooling, star formation, thermal stellar feedback and chemical enrichment from Type II SN, Type Ia SN and mass return from stellar winds.

GASOLINE contains metal cooling based on radiative transfer found in CLOUDY as described in Shen et al. (2010). The CLOUDY cooling was calculated using an external UV radiation field starting at  $z = 8.9$ . Star formation is calculated and SN feedback is implemented using the blastwave formalism as described in Stinson et al. (2006). Recent examples of simulations that use similar physics include Governato et al. (2007, 2009, 2010), Stinson et al. (2010), Brooks et al. (2011) and Guedes et al. (2011).

Star formation is calculated using the commonly used recipe described in Katz (1992). Stars form from gas below a maximum temperature of  $15\,000\text{ K}$  and above a density of  $1\text{ cm}^{-3}$  with an efficiency of 5 per cent. However, the high-resolution runs presented in Governato et al. (2010) and Guedes et al. (2011) used a higher threshold for star formation ( $100$  and  $5\text{ cm}^{-3}$ , respectively) that leads to more efficient gas outflows than the SF recipe adopted for the Aquila simulation. The star particles are treated as single stellar populations using the IMF described in Kroupa, Tout & Gilmore (1993). In this context, ejecta from both Type II and Type Ia SNe are considered. These SNe feed both energy and metals back into the interstellar medium gas surrounding the region where they formed. Type II

SNe deposit  $7 \times 10^{50}$  erg of energy into the surrounding interstellar medium. Since this gas is dense, it would be quickly radiated away due to the efficient cooling. For this reason, cooling is disabled for particles inside the blast region  $r_{\text{CSO}} = 10^{1.74} E_{51}^{0.32} n_0^{-0.16} \bar{P}_{04}^{-0.20}$  pc for the length of time  $t_{\text{CSO}} = 10^{6.85} E_{51}^{0.32} n_0^{0.34} \bar{P}_{04}^{-0.70}$  yr given in McKee & Ostriker (1977).

In summary, the input parameters used for the simulations analysed in this paper are  $n_{\text{th}} = 1.0 \text{ cm}^{-3}$ ,  $c_* = 0.05$  and  $E_{\text{SN}} = 0.7 \times 10^{51}$  erg.

#### A8 The RAMSES models (R, R-LSFE, R-AGN)

The RAMSES code is an Eulerian AMR code that uses the particle mesh techniques for the  $N$ -body part (stars and dark matter) and a shock-capturing, unsplit second-order MUSCL scheme for the fluid component. For the latter, we use the MinMod slope limiter and the HLLC Riemann solver, ensuring both stability and proper capturing of discontinuities (Teyssier 2002).

Star formation is implemented stochastically using a Schmidt law, with density threshold for star formation held fixed at  $n_{\text{th}} = 0.1 \text{ cm}^{-3}$  and an efficiency parameter chosen between 1 (R-LSFE) and 5 per cent (R, R-AGN). Stellar feedback is modelled using a thermal dump of  $10^{51} \text{ erg SN}^{-1}$ , assuming a Salpeter IMF. Cooling is performed using a tabulated cooling function depending on gas metallicity, the latter being modelled self-consistently as a additional scalar hydro variable and initiated during SNe explosions with a yield  $y = 10$  per cent (Rasera & Teyssier 2006; Dubois & Teyssier 2008).

The grid is refined following a quasi-Lagrangian strategy, each cell being refined if the number of dark matter particles exceeds eight, or if the baryonic mass (star + gas) exceeds eight times the initial mass resolution set by the Aquila initial conditions. In order to avoid catastrophic refinement at early times, we carefully trigger new levels so that the maximum level *levelmax* is opened only at expansion factor  $a = 0.8$ , the previous one at  $a = 0.4$ , *levelmax*-2 at  $a = 0.2$  etc. This ensures that the resolution of the grid in physical units remains roughly constant (although in a discrete, stepwise sense). For the level-6 simulation, we have set *levelmax* = 17 (cell size is 1 kpc physical) and for the level-5 simulation, *levelmax* = 19 (cell size is 261 pc physical). This corresponds also to the maximum levels reached by a pure dark matter simulation with the same number of dark matter particles for each case, minimizing spurious effects due to two-body relaxation. For the R-AGN simulation only, AGN feedback has been implemented using the model of Booth & Schaye (2010); see Teyssier et al. (2011) for details. A sphere of size four cells is defined around each sink particle defining the SMBH. This spherical region is used to determine the accretion rate on the SMBH, but also to spread on the grid the AGN feedback energy, using only a pure thermal dump.

To summarize, the RAMSES simulations use a star formation density threshold of  $n_{\text{th}} = 0.1 \text{ cm}^{-3}$ , an energy per SN of  $10^{51}$  erg and a star formation efficiency of 5 (R and R-AGN) or 1 per cent (R-LSFE).

#### A9 The AREPO model (AREPO)

The AREPO code (Springel 2010a) is a novel pseudo-Lagrangian hydrodynamical code that works with an unstructured, fully dynamic and adaptive mesh. The mesh is defined as the Voronoi tessellation (e.g. Okabe 2000) of the simulation volume generated by a set of mesh-generating points. The hydrodynamics is calculated with a second-order accurate finite volume approach on this mesh, based on the MUSCL-Hancock scheme that is also widely employed in

ordinary Eulerian mesh codes. This involves a spatial reconstruction step and flux estimates at all cell faces by solving Riemann problems at the interfaces. The very important new ingredient in the scheme is that the mesh-generating points are allowed to move freely during the calculation. In particular, they can be moved with the local fluid velocity such that the mesh dynamically follows the fluid motion without showing any problematic mesh twisting effects. In this mode, AREPO minimizes advection errors in the hydrodynamics and produces Galilean-invariant results that help to avoid accuracy problems with high-velocity cold flows that can occur in ordinary mesh codes. A more detailed description and an investigation of numerous test problems that demonstrate the high accuracy of the method can be found in Springel (2010a).

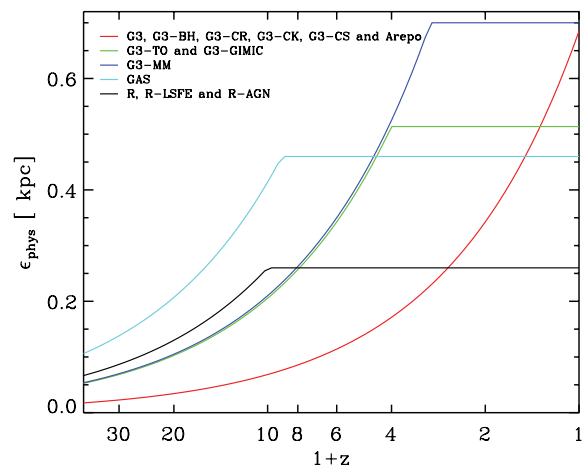
Even though the hydrodynamics is solved completely differently in AREPO than in GADGET3, the Lagrangian character of both codes makes it possible to implement the physics of star formation and feedback in very similar ways in both codes. In fact, the AREPO simulations analysed here implement exactly the same subgrid model for dense gas as well as the same SN feedback recipe as our default GADGET run g3. Since also the gravitational solver for both codes is identical, any differences found in the results hence reflect changes due to the numerical treatment of hydrodynamics alone.

## APPENDIX B: GRAVITATIONAL SOFTENING

Our 13 simulations have used a variety of choices for the evolution of the gravitational softening. This evolution is governed by two parameters:  $z_{\text{fix}}$  and  $\epsilon_g^{z=0}$ . The former divides the period where the softening changes from being fixed in comoving coordinates ( $z > z_{\text{fix}}$ ) to being fixed in physical coordinates ( $z \leq z_{\text{fix}}$ ). The different simulations have various choices for  $z_{\text{fix}}$ , as shown in shown in Table 2. The second parameter,  $\epsilon_g^{z=0}$ , is the value of the gravitational softening at the present time, and also varies slightly for our 13 simulation (Table 2). Fig. B1 shows the evolution of the gravitational softening in physical coordinates for our simulations.

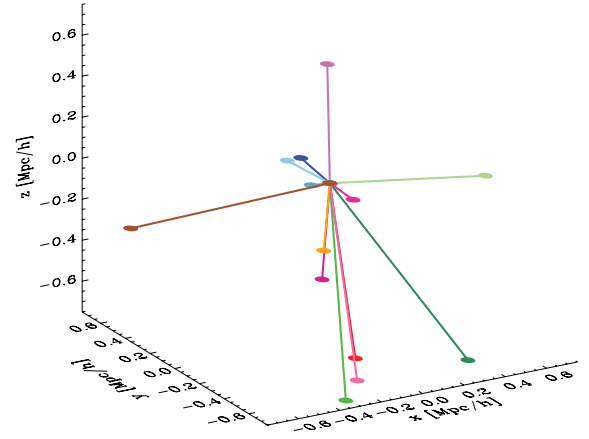
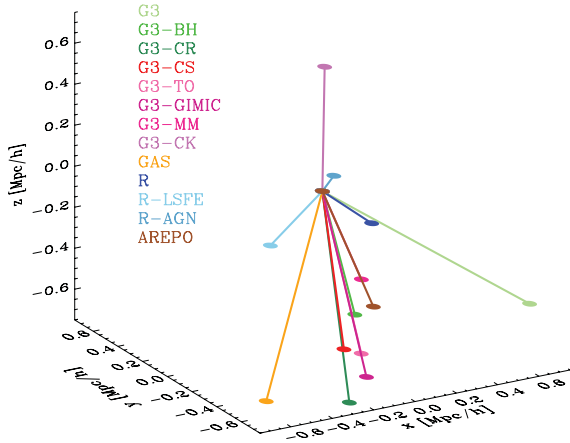
## APPENDIX C: DISC ORIENTATION

Fig. C1 illustrates, for our level-5 and level-6 runs, the spin of the stellar component of the simulated galaxies at  $z = 0$ . The two panels show, in a 3D rendering, the specific angular momentum vector of all stars in each galaxy, normalized to the maximum among all



**Figure B1.** Evolution of the *physical* gravitational softening in the different models.





**Figure C1.** Disc orientation of the level-5 (left-hand panel) and level-6 (right-hand panel) resolution simulations. The vectors indicate the orientation of the angular momentum of galactic stars, and are normalized to the maximum among all simulations.

simulations. Clearly, discs are not all aligned in the same direction, nor is the specific angular momentum the same. This is not surprising, given the wide range of stellar masses spanned by the different simulations. In spite of this, the orientation is actually similar for some simulations. As explained in Section 2.4, for orientation-dependent diagnostics we have rotated each simulated galaxy to a

new coordinate system where the angular momentum vector of its stellar component coincides with the  $z$  direction.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.